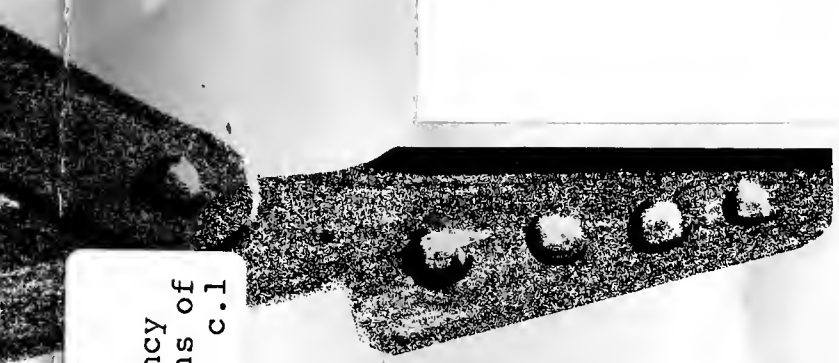
Robotics Research Technical Report

Review of Multifrequency Channel Decompositions of Images and Wavelet Models

by

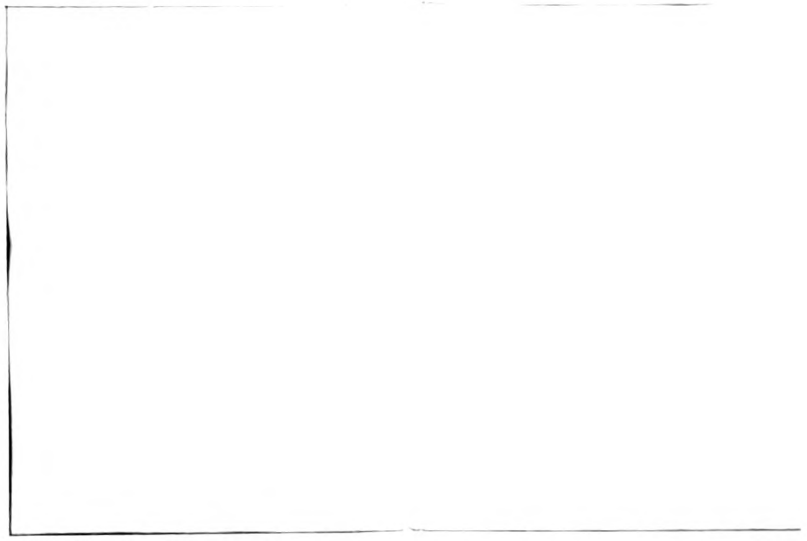
Stephane G. Mallat

Technical Report No. 412
Robotics Report No. 178
November, 1988



New York University
nt Institute of Mathematical Sciences

Computer Science Division
251 Mercer Street New York, N.Y. 10012



Review of Multifrequency Channel Decompositions of Images and Wavelet Models

by

Stephane G. Mallat

Technical Report No. 412
Robotics Report No. 178
November, 1988

New York University
Dept. of Computer Science
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, New York 10012

Work on this paper has been supported by DARPA/Office of Naval Research Grant N00014-85-K-0807.

REVIEW OF MULTIFREQUENCY CHANNEL DECOMPOSITIONS OF IMAGES AND WAVELET MODELS

Stephane G. Mallat

Dept. of Computer Sciences
Courant Institute of Mathematical Sciences
New York University

ABSTRACT

In this paper we review the recent multichannel models developed in psycho-physiology, computer vision and image processing. In psycho-physiology, multichannel models have been particularly successful for explaining some low-level processing in the visual cortex. The expansion of a function in several frequency channels provides a representation which is intermediate between a spatial and a Fourier representation. We describe the mathematical properties of such decompositions and introduce the wavelet transform. We review the classical multiresolution pyramidal transforms developed in computer vision and show how they relate to the decomposition of an image in a wavelet orthonormal basis. In the last section we discuss the properties of the zero-crossings of multifrequency channels. Zero-crossings representations are particularly well adapted for pattern recognition in computer vision.

1. Introduction

Within the last 10 years, multi-frequency channel decompositions have found many applications in image processing. In psycho-physiology of human vision, multichannel models have also been particularly successful for explaining some low-level biological processings. The

expansion of a function in several frequency channels provides a representation which is intermediate between a spatial and a Fourier representation. In functional analysis, this kind of transform appeared with the work of Little-Wood Payley in the 1930's . More research has recently been focused on this domain with the modelization of a new decomposition called the wavelet transform. In this paper we review the recent multichannel models developed in psycho-physiology, computer vision and image processing. We motivate the models within each of these disciplines and show how they relate to the wavelet transform.

In psychophysics and physiology of human vision, many evidences have been gathered showing that the retina image is decomposed into several spatially oriented frequency channels. In the first section of this paper, we describe the experimental motivations of this model. Biological studies of human vision have always been a source of ideas for computer vision and image processing research. Indeed, the human visual system is generally considered as an optimal image processor. The goal is not to imitate the processings implemented in the human brains but rather to understand the motivations of such processings and analyze their applications to computer vision problems. From this point of view, the recent experimental findings in psychophysics and physiology open challenging questions. In order to get a better understanding of multichannel decompositions, we review the main mathematical results in this domain. The best know decomposition which is intermediate between a spatial and a frequency representation is the window Fourier transform. We describe its properties but also show from a mathematical point of view why it is not convenient to analyze images. The wavelet transform was introduced by J. Morlet to compensate the inconveniences of a window Fourier transform. It is computed by expanding the signal on a family of functions which are the dilate and translate of a unique function $\psi(x)$. A. Grossmann and J. Morlet [22] have shown that any function of $\mathbf{L}^2(\mathbf{R})$ can be characterized from its decomposition on the wavelet family $\left[\sqrt{s}\ \psi(s(x-u))\right]_{(s,u)\in \mathbf{R}^2}$. A wavelet transform can be interpreted as a decomposition into a set of frequency channels having the same bandwidth on a logarithmic scale. We review the most important properties of a wavelet

transform and describe its discretization as studied by I. Daubechies [12]. A very important particular case of discrete wavelet transform was found by Y. Meyer [45] and J. Stromberg [54]. They proved that there exist some wavelets $\psi(x)$ such that $\left[\sqrt{2^j}\,\psi(2^j(x-2^{-j}n))\right]_{(j,n)\in Z^2}$ is an orthonormal basis of $L^2(R)$. Wavelet orthonormal bases provide an important new tool in functional analysis. Indeed, it was believed that we could not build simple orthonormal bases of $L^2(R)$ whose elements had a good localization both in the spatial and Fourier domains. These bases have already found many applications in pure and applied mathematics [28, 34, 56], in quantum mechanics [16, 47] and signal processing [31].

In computer vision, multifrequency channel decompositions are interpreted through the concept of multiresolution. Generally, the structures that we want to recognize have very different sizes. Hence, it is not possible to define a priori an optimal resolution for analyzing images. Several researchers [24, 42, 51] have developed pattern matching algorithms which process the image at different resolutions. Some pyramidal implementations have been developed for computing these decompositions [4, 10, 49]. A multiresolution transform decomposes the signal in a set of frequency channels of constant bandwidth on a logarithmic scale. It can be interpreted as a discrete wavelet transform. We review the wavelet multiresolution model [38] which provides a mathematical interpretation of the concept of resolution. We see in particular that a large class of wavelet orthonormal bases can be computed from quadrature mirror filters [39].

Multifrequency channel decompositions are well adapted for data compression in image coding. We show that this efficiency is due to the intrinsic statistical properties of images and to the ability of such representations to match the sensitivity of human vision. For pattern recognition applications, it is also necessary to build a signal representation which translates when the signal translates. Indeed, the representation of a pattern should not depend upon its position. When a pattern is translated, its representation should be translated without being modified. The pyramidal multiresolution representations as well as discrete wavelet transforms do not have this translation property. In the last paragraph, we study the properties of representations based on

zero-crossings of multifrequency channels. These representations do translate and for a particular class of band-pass filters, the zero-crossings provide the location of the signal edges. It remains to show that a zero-crossing representation can provide a complete and stable signal decomposition. We review previous results on zero-crossings properties and explain how the problem can be expressed through the wavelet model.

2. Multichannel models in Psychophysics and Physiology of vision

In this paragraph, we summarize some experimental results showing that multifrequency channel decompositions seems to be implemented in the human visual cortex. For further details, we refer to tutorials by M. Georgeson [20], and M. Levine [35]. Over the past 20 years, a large effort has been devoted in psychophysics and physiology to analyze the response of the human visual system to stimuli having particular orientation and frequency tunings. Linear models have been partly successful to explain some experimental data. The simplest which was first developed in psychophysiology consists in considering the human visual system as a linear filter. Fig. 1 illustrates the anatomical pathway in the human visual system. Photoreceptors in the eyes measure the light input intensity. This information is processed by bipolar and ganglion cells in the retina and is transmitted through the optic nerve. The optic nerve ends in a relay station (lateral geniculate nucleus) whose axons project to the visual cortex. Replacing these different stages by a global linear filter is clearly an extremely simplified model but it gives some insights about the visual system sensitivity. Given this hypothesis, Campbell and Green [6] tried to measure the global transfer function of the visual system. In their experiments, the visual stimulus shown to the observer were vertical sinusoidal gratings of different spatial frequencies (see Fig. 2). In psychophysics, frequencies are measured in cycles per degree of visual angle subtented by the eye. The transfer function $H(\omega)$ of the visual system is defined as the ratio of the contrast perceived by the observer to the real contrast of the stimulus. The contrast is given by

$$C = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} ,$$

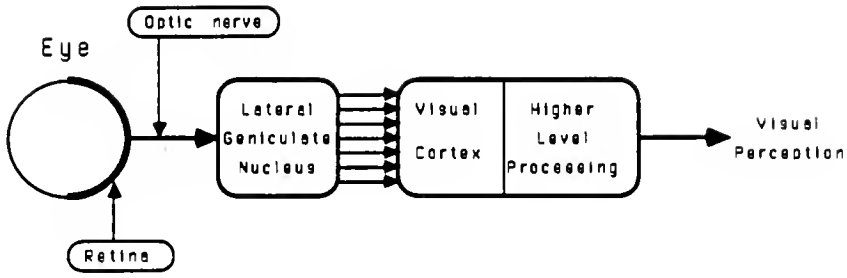


Fig. 1 : *Illustration of the anatomical visual pathway. The higher level processings are the least understood and are difficult to evaluate in psychophysical experiments.*

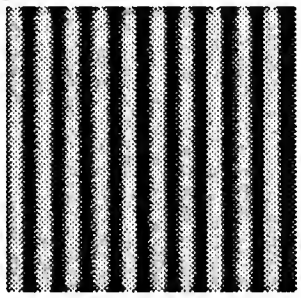


Fig. 2 : *This image is a typical visual stimulus used in psychological experiments for computing the transfer function of the visual system. It consists of sinuoidal gratings whose frequency varies along the experiments. In order to evaluate the sensitivity to orientation, these gratings are rotated.*

where L_{max} and L_{min} are the maximum and minimum luminance of the stimuli. In order to estimate this transfer function, a solution which is widely adopted is to measure the *Contrast Sensitivity Function*. At each frequency ω , we measure the minimum contrast $C_t(\omega)$ necessary to distinguish the sinusoidal gratings from a uniform background. This contrast is called the contrast threshold. The contrast sensitivity function is then defined by

$$CSF(\omega) = \frac{1}{C_t(\omega)}, \quad \text{and} \quad H(\omega) = CSF(\omega) .$$

Many experiments [5, 6, 32] have been made to measure the function $CSF(\omega)$ and they agree approximatively with the function shown in Fig. 3. Although this linear model is clearly over simplified, it shows qualitatively the sensitivity of the human system to stimuli of different frequencies.

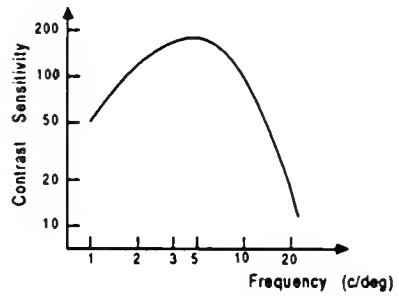


Fig. 3 : *Contrast Sensitivity Function (redrawn from Kulikowski and King-Smith [32]). The visual system has a maximum sensitivity to contrast when the frequency of the stimulus is around 5 c/deg.*

With further experiments, F. Campbell and J. Robson [8] have shown that a better model is to consider that the image retina is processed through independent frequency channels. These experiments were based on adaptation technics. If a stimulus is shown to an observer for a long time, the visual sensitivity for the same kind of stimuli decreases. This behavior is called an adaptation process. F. Campbell and J. Robson [8] have shown that if the visual system adapts to a sinusoidal grating of a given frequency ω_0, the sensitivity decreases for any stimuli whose frequency is in a frequency band around ω_0. However, outside this frequency band, the sensitivity is not perturbated. These experiments indicate that at some stage, the visual information is probably processed through independent frequency channels. In order to simplify their analysis, F. Campbell and J. Robson supposed that the retina image was decomposed through independent

band-pass linear filters as shown in Fig. 4. Their first estimation of the frequency bandwidth of these filters was very narrow. However, other experiments by M. Georgeson [19] and Nachmias [46] have then contradicted their results. They showed that the frequency bandwidth of these filters is more likely to be around an octave.

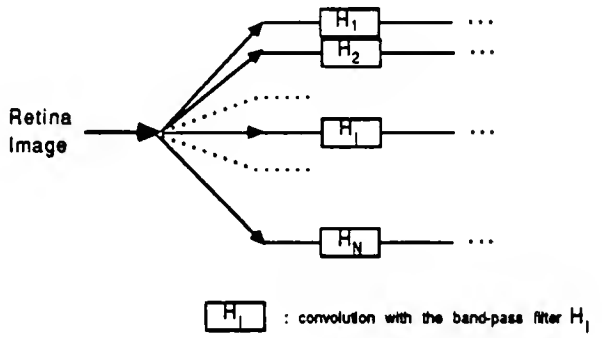


Fig. 4 : *Multichannel model. The retina image is supposed to be filtered by independent band-pass filters. These filters have approximatively the same band-width on a logarithmic scale.*

Other psychophysical experiments have shown that the visual sensitivity to a sinusoidal grating also depends upon its spatial orientation. The results of Campbell and Kulikowski [7] show that the human visual system has a maximum sensitivity when the signal has an orientation of 0° or 90°. In between, the sensitivity decreases monotonically reaching a minimum at 45°. The filters of the model shown in Fig. 4 must therefore include a spatial orientation selectivity.

This filter bank model only provides a qualitative description of some low-level processing of the visual system. In particular it does not take into account the nonlinearities of the biological processing. However, recent physiological experiments gave some support to such approaches. Cell recordings experiments are generally performed on cats and monkeys which have a visual cortex organization similar to the human one. In the cats visual cortex, Hubel and Wiesel [25] discovered a class of cells whose response depend upon the frequency and orientation of the visual stimuli. These cells are called simple cells. Maffei and Fiorentini [36] have shown that

their response is reasonably linear and that they can modelized with linear filters. Several groups of researcher have recorded their corresponding impulse responses [2, 37, 57]. These studies showed that the bandwidths of simple cells range from 0.6 to 2.0 octaves with an average of 1.3 octave. They have a receptive field of varying size depending on their frequency tuning [48]. The response of simple cells also depends upon the spatial orientation of the stimuli. M. Webster and R. De Valois [59] showed that for a given cell, this orientation selectivity is independent from the frequency of the stimulus (see Fig. 5). The impulse response of simple cells have been modeled by J. Daugmann [13, 14] with Gaussians modulated by sinusoidal waves. As explained in the next paragraph, these functions functions generate a particular window Fourier transform called Gabor transform. Fig. 5 shows the comparison between the impulse response of a simple cell and the corresponding Gabor function model. These graphs clearly show that simple cells behave like band pass filter with a spatial orientation tuning. The support of the transfer function a cell is called the receptive field. It corresponds to the domain of the retina where the input image influences the firing of the cell.

Many evidences have now been gathered around this multifrequency channel modeling of some of the low-level visual cortex processing. However, there is no real knowledge of what type on information is extracted from this decomposition and how it relates to further processing by complex and hyper-complex cells [48]. Since we know that the human visual cortex is an excellent image processor, this low-level biological model raises important questions from an image processing point of view. What is the advantage of decomposing a signal into several frequency channels ? Is it related to the intrinsic statistical properties of images ? Does it lead to a better reorganization of the image information ? If we do accept that such a decomposition offers an interesting representation of images, it remains to find out how to process these different frequency channels. What type of information do we want to extract ? Should we process each channel independently or compare the values of the signal within each band ? In the following paragraphs, we show that some results in mathematics, computer vision and image coding give

elements of answers to these questions. Our primary goal is not to build a model of the human visual cortex but rather to justify the use of such decompositions in image processing.

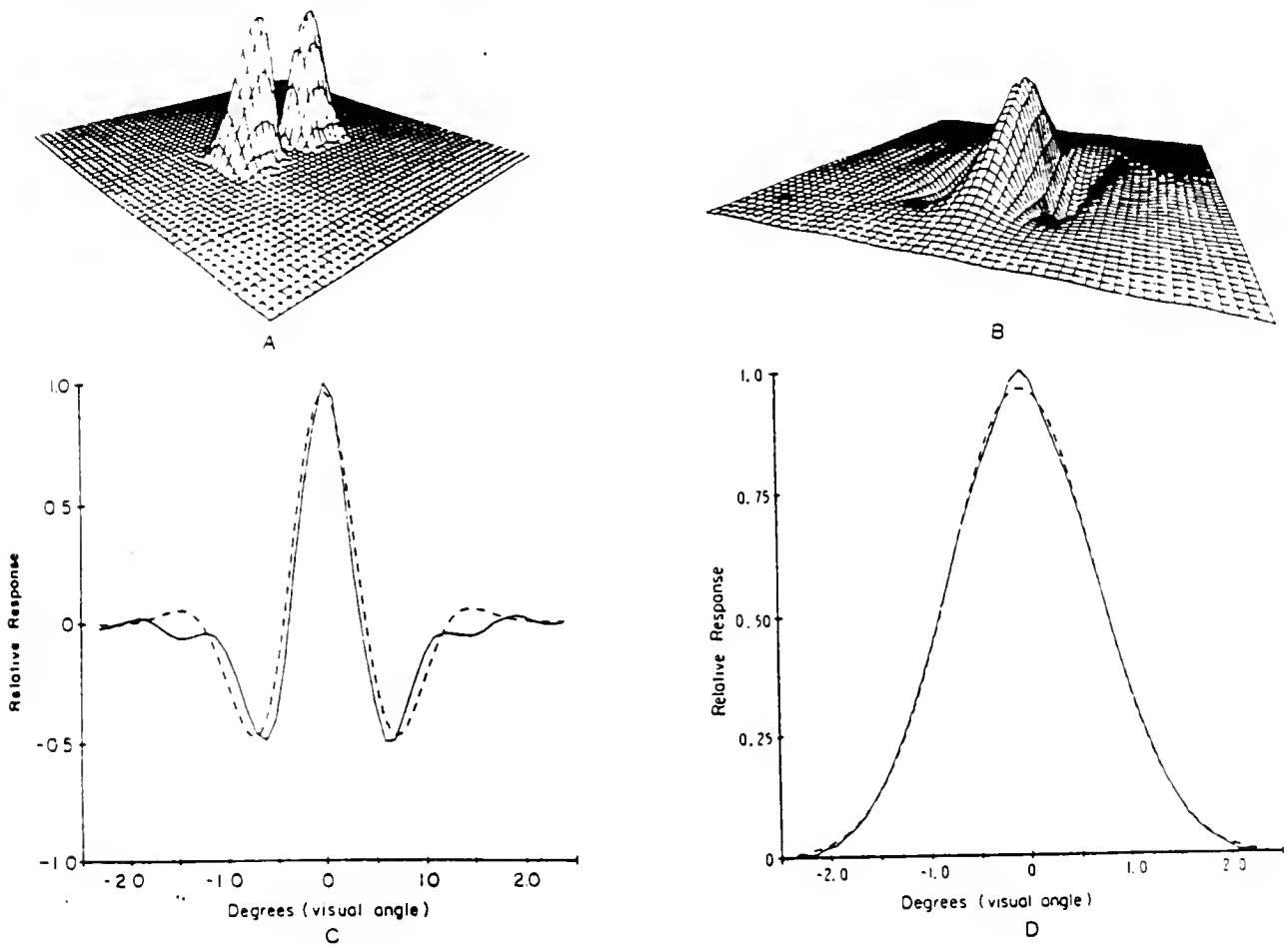


Fig. 5 : *(Reprint from M. Webster and R. De Valois [59]). A: This surface is the two-dimensional transfer function of a simple cell. It is a band-pass oriented filter. Its band-width is of 0.94 octaves. B: Impulse response computed by taking the inverse Fourier transform of Fig. A. C and D: Cross sections of the impulse response respectively along the x and y axes. The dashed lines gives the best fitting Gabor functions.*

3. Mathematical analysis of multichannel models

In this paragraph we review the mathematical properties of multifrequency channel decompositions. No proof are given but we refer to original works. Most of the results are first introduced for one-dimensional functions and then generalized in two dimensions if needed. We review the properties of a window Fourier transform which is the more classical intermediate decomposition between the spatial and Fourier representation. We also describe the inconveniences of this decomposition for analyzing signals like images. The wavelet transform is then introduced with respect to a window Fourier transform. More details can be found in a complete article by I. Daubechies [12] and an advance functional analysis book by Y.Meyer [44].

Notation

$\mathbf{Z}$, $\mathbf{R}$ and $\mathbf{R}^+$ denote respectively the sets of integers, real numbers and positive real numbers. $\mathbf{L}^2(\mathbf{R})$ denotes the Hilbert space of measurable, square-integrable one-dimensional functions f(x) having a complex value. We suppose that our signals are finite energy functions $f(x) \in \mathbf{L}^2(\mathbf{R})$. For a pair of functions $f(x) \in \mathbf{L}^2(\mathbf{R})$, $g(x) \in \mathbf{L}^2(\mathbf{R})$, the inner product of f(x) with g(x) is written

$$< g(x) , f(x) > = \int_{-\infty}^{+\infty} g(x) \overline{f(x)} \, dx \quad , \tag{1}$$

where $\overline{f(x)}$ is the complex conjugate of f(x) . The norm of f(x) in $\mathbf{L}^2(\mathbf{R})$ is given by

$$||f||^2 = \int_{-\infty}^{+\infty} |f(x)|^2 \, dx \quad . \tag{2}$$

We denote the convolution of two functions $f(x) \in \mathbf{L}^2(\mathbf{R})$ and $g(x) \in \mathbf{L}^2(\mathbf{R})$ by

$$f * g(u) = (f(x) * g(x))(u) = \int_{-\infty}^{+\infty} f(x) \, g(u-x) \, du \quad . \tag{3}$$

The dilation of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ by a scaling factor s is written

$$f_s(x) = \sqrt{s} \, f(sx) \quad . \tag{4}$$

The reflection of f(x) about 0 is written

$$\check{f}(x) = f(-x) \ . \tag{5}$$

The Fourier transform of $f(x) \in \mathbf{L}^2(\mathbf{R})$ is written $\hat{f}(\omega)$ and is defined by

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)\, e^{-i\omega x}\, dx \ . \tag{6}$$

3.1. Definition of a window Fourier transform

From the Fourier transform of a function f(x), we get a measure of the irregularities (high frequencies) but this information is not spatially localized. Indeed, the Fourier transform $\hat{f}(\omega)$ is defined through an integral which covers the whole spatial domain. It is therefore difficult to find the position of the irregularities. In order to localize the information provided by the Fourier transform, Gabor [17] defined a new decomposition using a spatial window $g(x)$ in the Fourier integral. This window is translated along the spatial axis in order to cover the whole signal. At a position u and for a frequency ω, the window Fourier transform of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ is defined by :

$$Gf(\omega,u) = \int_{-\infty}^{+\infty} e^{-i\omega x}\, g(x-u)\, f(x)\ dx \ . \tag{7}$$

It measures locally, around the point u, the amplitude of the sinusoidal wave component of frequency ω. In the original Gabor transform, the window function g(x) is a Gaussian. It has since been generalized for any type of window function and is called a window Fourier transform [29]. The window function is generally a real even function and the energy of its Fourier transform is concentrated in the low-frequencies (see Fig. 6). It can be viewed as a low-pass filter. For normalization purposes, we suppose that the energy of $g(x)$ is equal to 1 :

$$\|g\|^2 = \int_{-\infty}^{+\infty} |g(x)|^2\, dx = 1 \ .$$

Let us denote

$$g_{\omega_0,u_0}(x) = e^{i\omega_0 x}\, g(x-u_0) \ .$$

A window Fourier transform can also be interpreted as the inner products of the function f(x) with the family of functions $\left\{ g_{\omega,u}(x) \right\}_{(\omega,u)\in \mathbf{R}^2}$:

$$Gf(\omega,u) = < f(x) , g_{\omega,u}(x) > \ . \tag{8}$$

In quantum physics, such a family of functions is called a family of coherent states. The Fourier transform of $g_{\omega_0,u_0}(x)$ is given by

$$\hat{g}_{\omega_0,u_0}(\omega) = e^{-iu_0\omega}\, \hat{g}(\omega-\omega_0) \ , \tag{9}$$

where $\hat{g}(\omega)$ is the Fourier transform of g(x) . A family of coherent states thus corresponds to a translation in the spatial domain (parameter u) and in the frequency domain (parameter ω) of the function g(x) . This double translation is represented in a phase-space where one of the axis corresponds to the spatial parameter u and the other to the frequency parameter ω (see Fig. 7).

Let us now describe how a window Fourier transform relates to a spatial or a frequency representation. Let σ_u be the standard deviation of g(x) :

$$\sigma_u^2 = \int_{-\infty}^{+\infty} x^2\, |g(x)|^2 \, dx \ . \tag{10}$$

Let σ_ω be the standard deviation of the Fourier transform of g(x) :

$$\sigma_\omega^2 = \int_{-\infty}^{+\infty} \omega^2\, |\hat{g}(\omega)|^2 \, d\omega \ . \tag{11}$$

The function $g_{\omega_0,u_0}(x)$ is centered in u_0 and has a standard deviation of σ_u in the spatial domain. Its Fourier transform given by equation (9) is centered in ω_0 and has a standard deviation of σ_ω. By applying the Parseval theorem on equation (8) we get

$$Gf(\omega_0,u_0) = \int_{-\infty}^{+\infty} f(x)\, \overline{g_{\omega_0,u_0}(x)}\, dx = \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \overline{\hat{g}_{\omega_0,u_0}(\omega)}\, d\omega \ . \tag{12}$$

This equation shows that in the spatial domain $Gf(\omega_0,u_0)$ provides a description of f(x) around

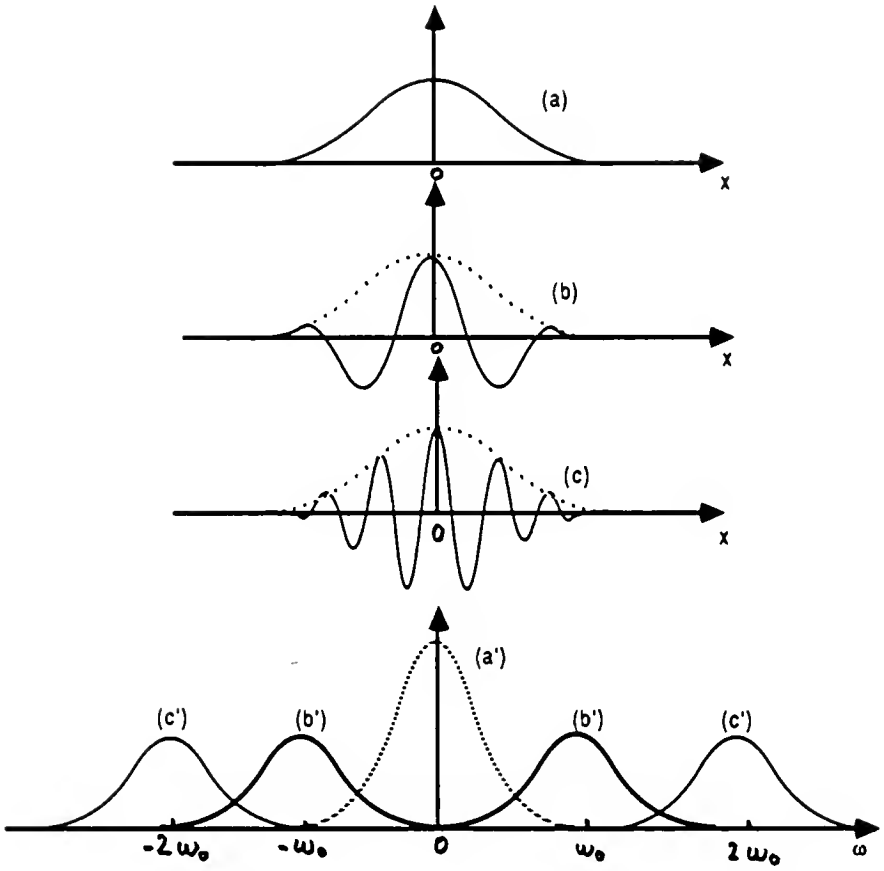


Fig. 6 : *(a): Window function* $g(x)$. *(b): Graph of* $g(x)\cos(\omega_0 x)$. *(c): Graph of* $g(x)\cos(2\omega_0 x)$. *All these curves have the same support but the number of cycles varies with the frequency of the sinusoidal modulation. The curves (a') , (b') and (c') are respectively the Fourier transform of* $g(x)$, $g(x)\cos(\omega_0 x)$ *and* $g(x)\cos(2\omega_0 x)$. *They have the same bandwidth but have different positions on the frequency axis.*

u_0 within a standard deviation of σ_u. In the Fourier domain is describes the behavior of $\hat{f}(\omega)$ around ω_0 within a standard deviation of σ_ω. This can be represented in the phase-space by the resolution cell $[u_0 - \sigma_u , u_0 + \sigma_u] \times [\omega_0 - \sigma_\omega , \omega_0 + \sigma_\omega]$ as shown in Fig. 6. The surface and shape of the resolution cell is independent from u_0 and ω_0. The uncertitude principle implies

$$\sigma_u^2 \, \sigma_\omega^2 \geq \frac{\pi}{2} \quad . \tag{13}$$

The resolution cell can therefore not be smaller than $2\sqrt{2\pi}$. The uncertitude inequality reaches its upper limit if and only if g(x) is a Gaussian. Hence, the resolution in the phase-space is maximum when the window function is a Gaussian.

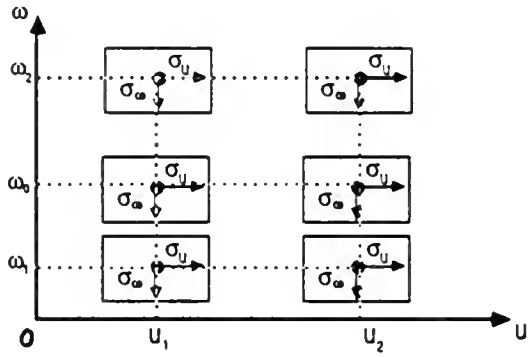


Fig. 7 : *Phase-space representation. The vertical axis gives the frequency* ω *whereas the horizontal axis gives the spatial position* u. *A window Fourier coefficient* $Gf(\omega_0,u_0)$ *provides a description of* $f(x)$ *within the resolution cell* $[u_0-\sigma_u\ ,\ u_0+\sigma_u]\times[\omega_0-\sigma_\omega\ ,\ \omega_0+\sigma_\omega]$.

3.2. Properties of a window Fourier transform

One can show that a window Fourier transform is an isometry (to a proportionality coefficient) from $\mathbf{L}^2(\mathbf{R})$ into $\mathbf{L}^2(\mathbf{R}^2)$:

$$\int_{-\infty}^{+\infty} |f(x)|^2\,dx \;=\; \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |Gf(\omega,u)|^2\;d\omega\,du \quad . \tag{14}$$

Hence, the reconstruction of f(x) from $Gf(\omega,u)$ is stable and

$$f(x) \;=\; \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} Gf(\omega,u)\,g(u-x)\,e^{i\omega x}\;d\omega\,du \quad . \tag{15}$$

A window Fourier transform is a redundant representation. We can define a complete and stable representation by sampling uniformly the spatial parameter u and the frequency parameter ω. Let u_0 and ω_0 be the sampling intervals in both domains. A discrete Fourier transform is

defined by

$$\forall n \in \mathbf{Z} \ , \ \forall m \in \mathbf{Z} \quad G_d f(m,n) = \int_{-\infty}^{+\infty} e^{-im\omega_0 x} \, g(x-nu_0) \, f(x) \, dx \ . \tag{16}$$

This discretization corresponds to a uniform sampling of the phase-space as shown in Fig. 8. A discrete window Fourier transform is equivalent to a division of the frequency axis into intervals separated by ω_0 (see Fig. 6). In each of these intervals, the signal is sampled at a rate $\frac{1}{u_0}$. I.Daubechies [12] made of a thorough study of the completeness and stability of a discrete window Fourier transform. Intuitively, the sampling intervals u_0 and ω_0 must be chosen in order to cover the whole phase space with the resolution cells shown in Fig. 7. Formally, to reconstruct any function $f(x) \in \mathbf{L}^2(\mathbf{R})$ from the set of sample $\left[G_d f(n,m) \right]_{(n,m) \in Z^2}$, the operator

$$\mathbf{L}^2(\mathbf{R}) \xrightarrow{\quad G_d \quad} \mathbf{l}^2(\mathbf{Z}^2)$$

must be invertible on its range and have a bounded inverse. Each sample $G_d f(n,m)$ can also be expressed as

$$G_d f(m,n) = < f(x) \, , \, e^{im\omega_0 x} \, g(x-nu_0) \, f(x) > \ = \ < f(x) \, , \, g_{m\omega_0, nu_0}(x) > \ . \tag{17}$$

The properties of a discrete Fourier transform thus depend upon the family of functions $\left[g_{m\omega_0, nu_0}(x) \right]_{(n,m) \in Z^2}$. In order to invert G_d, I. Daubechies [12] has shown that ω_0 and u_0 must verify

$$\omega_0 \, u_0 \leq 2\pi \ . \tag{18}$$

When $\omega_0 \, u_0 = 2\pi$, we reach the Nyquist frequency limit.

Although several researchers have tried to modelize the impulse response of simple cells with Gabor functions, it is unlikely that the human visual cortex implements some type of window Fourier transform. Indeed, we saw that a window Fourier transform decomposes a function in a set of frequency intervals having the same size. On the opposite, experimental data indicate

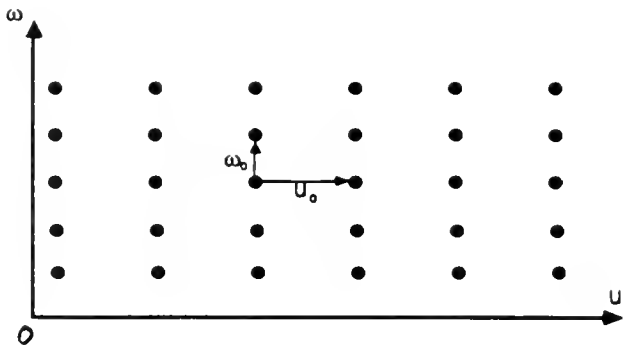


Fig. 8 : *Sampling pattern of a discrete window Fourier transform in the phase space. Since the resolution cells are identical everywhere in the phase-space, the sampling is uniform.*

that the retina image is decomposed in a set of frequency channels having approximatively a constant bandwidth on a logarithmic scale (octave). The measured impulse responses of simple cells do not have increasing number of cycles for a constant envelop like in a window Fourier transform (see Fig. 6). They rather have a support (receptive filed) of varying size.

Let us now study the application of a window Fourier transform for analyzing images. We saw that the spatial and frequency resolution of a window Fourier transform is constant. In the spatial domain, the information provided by this decomposition is unlocalized within intervals of size σ_u . It privileges a resolution of reference which depends upon the standard deviation σ_u of g(x) . If the signal has a discontinuity such as an edge, it is difficult with a window Fourier transform to locate this edge with a precision better than σ_u (see Fig. 9). This localization limit is generally not acceptable. If the signal has important features of very different sizes, we can not define an optimal resolution for analyzing the signal. This is typically the case with images. For example, in the image of a house, the pattern we want to analyze might range from the overall structure of the house to the drawings of one of the window curtains. This fixed resolution also introduces misleading high frequencies when decomposing local features. Let $e(x)$ be an edge

as shown in Fig. 9, and Δx be the variation step of this edge. The decomposition coefficients $G_d e(m,n) = < e(x) , e^{im\omega_0 x} g(x-nu_0) >$ are non-negligible even when $m\ \omega_0 \gg \frac{2\pi}{\Delta x}$. This is due to the difference of size between the support of g(x) and the variation step Δx . This numerical property makes it difficult to interpret the window Fourier coefficients when the features are very localized with respect to the size of the support of g(x) . For these reasons, a window Fourier transform is not convenient for analyzing images.

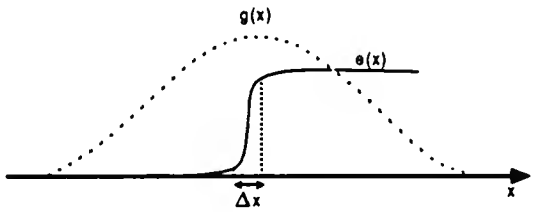


Fig. 9 : *With a window Fourier transform, a local feature such as an edge $e(x)$ can not be located with a precision better than the variance σ_u of the window function $g(x)$. The difference between the variation support Δx of the edge and σ_u also creates artificial high frequency components in the window Fourier transform decomposition.*

In order to avoid the inconveniences of a transform having a fixed resolution in the spatial and frequency domains, J. Morlet defined a new decomposition based on a dilation of the analyzing function. In the next paragraph, we describe the properties of this decomposition called the wavelet transform.

3.3. Definition of a wavelet transform

J. Morlet [22] defined the wavelet transform by decomposing the signal on a family of functions which are the translate and dilate of a unique function $\psi(x)$. The function $\psi(x)$ is called a wavelet and the corresponding wavelet family is given by $\left[\sqrt{s}\ \psi(s(x-u))\right]_{(s,u)\in \mathbf{R}^2}$.

The wavelet transform of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ is defined by

$$Wf(s,u) = \int_{-\infty}^{+\infty} f(x)\, \sqrt{s}\ \psi(s(x-u))\ dx \quad . \tag{19}$$

The idea of a wavelet decomposition is not totally new. It is very much related to some other types of space-frequency decompositions such as the Wigner-Ville transform. Some versions of the wavelet transform have been studied independently under other names like the scale-space decomposition of Witkin [60]. However, the formalization effort of J. Morlet and A. Grossmann [22] opened a broader field of applications to this approach and have led to important new mathematical results. Let us denote by

$$\psi_s(x) = \sqrt{s}\ \psi(sx)$$

the dilation of $\psi(x)$ by a factor s . A wavelet transform can be rewritten as inner products :

$$Wf(s,u) = < f(x)\, , \psi_s(x-u) > \quad . \tag{20}$$

It corresponds to a decomposition of f(x) on the family of functions $\left[\psi_s(x-u)\right]_{(s,u)\in \mathbf{R}^2}$. As shown in Fig. 10 , the functions $\psi_s(x)$ have the same shape than $\psi(x)$ but a support s times smaller. In the following, we suppose that the Fourier transform $\hat{\psi}(\omega)$ of $\psi(x)$ verifies the admissibility condition :

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\ d\omega \ < +\infty \quad . \tag{21}$$

This condition implies that $\hat{\psi}(0) = 0$ and that $\hat{\psi}(\omega)$ is small enought in the neighborhood of $\omega = 0$. The function $\psi(x)$ can thus be interpreted as a band-pass filter. For normalization purposes, we suppose that the energy of $\psi(x)$ is equal to 1 . Let us denote $\bar{\psi}_s(x) = \psi_s(-x)$. The wavelet transform at a point u and a scale s can also be written

$$Wf(s,u) = f * \bar{\psi}_s(u) \quad . \tag{22}$$

Hence, it corresponds to a filtering of f(x) with the band-pass filter $\bar{\psi}_s(x)$. The Fourier transform of $\psi_s(x)$ is given by

$$\hat{\psi}_s(\omega) = \frac{1}{\sqrt{s}} \hat{\psi}(\frac{\omega}{s}) .$$

On a logarithmic scale, the bandwidth of $\hat{\psi}_s(\omega)$ is independent of s . A wavelet transform decomposes the signal into a set of frequency bands having a constant size on a logarithmic scale (see Fig. 10).

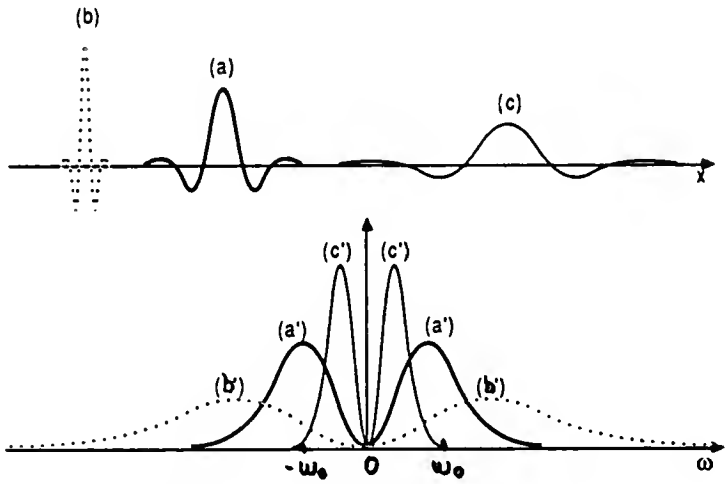


Fig. 10 : *(a): Graph of a wavelet* $\psi(x)$. *(b): Graph of* $\psi_{s_1}(x)$ *for* $s_1 > 1$. *(c): Graph of* $\psi_{s_2}(x)$ *for* $s_2 < 1$. *The curves (a') , (b') and (c') are respectively the Fourier transform of the function shown in (a) , (b) and (c). They have the same bandwidth on a logarithmic scale.*

In opposition to a window Fourier transform which has a fixed resolution in the spatial and frequency domains, the resolution of a wavelet transform varies with the scale parameter s . Let ω_0 be the center of the passing band of $\hat{\psi}(\omega)$ (see Fig. 10). Let σ_u and σ_ω be respectively the standard deviation of $\psi(x)$ in the spatial and frequency domains. One can easily show that the wavelet $\psi_s(x-u_0)$ has an energy concentrated around u_0 within a standard deviation of $\frac{\sigma_u}{s}$ in the spatial domain. In the frequency domain, its energy is concentrated around $s\omega_0$ within a standard deviation of $s\sigma_\omega$. In the phase-space, the resolution cell of this wavelet is therefore equal to $[u_0 - \frac{\sigma_u}{s} , u_0 + \frac{\sigma_u}{s}] \times [\omega_0 - s\sigma_\omega , \omega_0 + s\sigma_\omega]$. The surface of this resolution cell is

constant and equal to $4\,\sigma_u\,\sigma_\omega$ as for a window Fourier transform. However, depending on the scale s, the shape of the resolution cell varies. This is illustrated in Fig. 11. When the scale s is small, the resolution is coarse in the spatial domain and fine in the frequency domain. If the scale s increases, the resolution increases in the spatial domain and decreases in the frequency domain (see Fig. 11). In the next paragraph, we show that this variation of resolution enables the wavelet transform to zoom into the irregularities of the signal and characterize them locally.

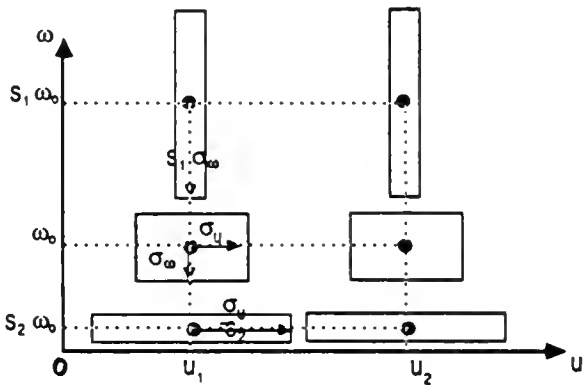


Fig. 11 : *In the phase-space, the shape of a wavelet resolution cell depends upon upon the scale. When the scale increases, the resolution increases in the spatial domain and decreases in the frequency domain. The surface of all the resolution cells is the same.*

The wavelet function $\psi(x)$ can either be real or complex. It can be useful to use a complex wavelet in order to separate the phase and modulus of the wavelet transform [22]. For this purpose, J. Morlet and A. Grossmann are using wavelets whose Fourier transform $\hat{\psi}(\omega)$ is equal to zero for $\omega < 0$. Such functions are called Hardy functions. The wavelet transform $Wf(s,u)$ is now a complex number. When the scale s is fixed and u varies, the function $Wf(s,u)$ is also a Hardy function. The phase and the modulus of the wavelet transform can easily be separated for any given scale s and position u. For some applications in signal processing, it can be particularly interesting to isolate the phase and the modulus [31].

Remark

There is a common misunderstanding in the psycho-physiological and computer vision literature around Gabor and wavelet transforms. A Gabor function is a Gaussian modulated by a sinusoidal wave. A Gabor function satisfies the admissibility condition (21) and can therefore be considered as a wavelet. If we build a transform based on a dilation of this function, it will be a wavelet transform and not a Gabor transform (window Fourier transform). Indeed, in order to define a Gabor transform, we must modify the frequency of the sinusoidal modulation without changing the size of the window function. This is much more than a terminology problem since the properties of a wavelet transform and a Gabor transform are very different.

3.4. Properties of a wavelet transform

J. Morlet and A. Grossmann [22] have shown that the wavelet transform W is an isometry (to a proportionality coefficient) from $\mathbf{L}^2(\mathbf{R})$ into $\mathbf{L}^2(\mathbf{R}^+ \times \mathbf{R})$:

$$\int_{-\infty}^{+\infty} \int_{0}^{+\infty} |Wf(s,u)|^2 \, ds \, du = C_\psi \int_{-\infty}^{+\infty} |f(x)|^2 \, dx \ , \tag{23}$$

where C_ψ is the constant defined in equation (21). From equation (23), we derive that the reconstruction of f(x) from $Wf(s,u)$ is stable and

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} Wf(s,u) \, \psi_s(x-u) \, ds \, du \ . \tag{24}$$

Similarly to a window Fourier transform, a wavelet transform is redundant. This redundancy can be expressed with an integrating kernel [23]:

$$Wf(s',u') = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} Wf(s,u) \, K(s,s',u,u') \, ds \, du \ , \tag{25}$$

$$\text{where} \quad K(s,s',u,u') = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \psi_s(x-u) \, \psi_{s'}(x-u') \, dx \ .$$

$K(s,s',u,u')$ is called a reproducing kernel. It expresses the redundancy between $Wf(s,u)$ and

$Wf(s',u')$. One can show that a function $F(s,u) \in \mathbf{L}^2(\mathbf{R}^+ \times \mathbf{R})$ is the wavelet transform of some function $f(x) \in \mathbf{L}^2(\mathbf{R})$ if and only if it satisfies equation (25) . The reproducing kernel equation is an important characterization of a wavelet transform.

The wavelet transform can be discretized by sampling both the scale parameter s and the translation parameter u . In order to build a complete representation, we must cover the phase-space with the resolution cells shown in Fig. 11. This can be done with an exponential sampling of the scale parameter. We therefore select a sequence of scales $(\alpha^n)_{n \in \mathbf{Z}}$, where α is the elementary dilation step. For each dilation parameter α^n , the function $\psi_{\alpha^n}(x)$ is a band-pass filter whose band-width is proportional to α^n . When u varies, $Wf(\alpha^n,u)$ is therefore a signal whose bandwidth is proportional to α^n . In order to characterize this signal, we must sample it uniformly at a rate proportional to α^n . Let $\frac{\alpha^n}{\beta}$ be the sampling rate at the scale α^n . The discrete wavelet transformed is defined by :

$$W_d f(m,n) = \int_{-\infty}^{+\infty} f(x)\, \psi_{\alpha^n}(x - \frac{m\beta}{\alpha^n})\, dx \quad . \tag{26}$$

Fig. 12 illustrates this sampling pattern in the phase-space. When the scale increases, the density of samples increases. Equation (26) can also be written as a uniform sampling of f(x) filtered by $\bar{\psi}_{\alpha^n}(x)$:

$$W_d f(m,n) = f * \bar{\psi}_{\alpha^n}(\frac{m\beta}{\alpha^n}) \quad . \tag{27}$$

On a logarithmic scale, a discrete wavelet transform can be interpreted as a decomposition in a set of frequency channels of constant size and separated by $log(\alpha)$. Along each channel the signal is sampled at the rate $\frac{\alpha^n}{\beta}$.

It is not possible to understand the properties of this transform by using the Nyquist theorem since the Fourier transform of $\psi(x)$ is not band-limited. With an approach similar to

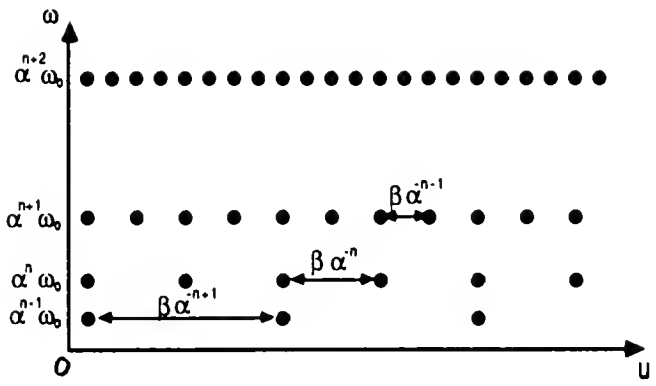


Fig. 12 : *Sampling of the phase-space corresponding to a discrete wavelet transform (adapted from I. Daubechies [12]). Each sample corresponds to an inner product with a particular wavelet.*

her study of the discrete window Fourier transform, I. Daubechies [12] analyzed the main properties of a discrete wavelet transform. She made a clear comparison of these two types of multichannel decompositions from a mathematical point of view. In order to reconstruct a function f(x) from the discrete wavelet transform $\left[W_d f(m,n)\right]_{(n,m)\in Z^2}$, the operator

$$\mathbf{L}^2(\mathbf{R}) \xrightarrow{\quad W_d \quad} l^2(\mathbf{Z}^2)$$

must be invertible on its range and have a bounded inverse. Since

$$W_d(m,n) = < f(x) , \psi_{\alpha^n}(x-\frac{m\beta}{\alpha^n}) > , \tag{28}$$

the properties of the operator W_d depend upon the family of functions $\left[\psi_{\alpha^n}(x-\frac{m\beta}{\alpha^n})\right]_{(n,m)\in Z^2}$. Let us suppose that $\psi(x)$ admits a derivative $\psi'(x)$ and that the decay at infinity of $\psi(x)$ and $\psi'(x)$ is given by

$$\psi(x) = O(x^{-2}) \quad \text{and} \quad \psi'(x) = O(x^{-2}) .$$

I. Daubechies [12] showed that if there exists two constant $C_1 > 0$ and $C_2 > 0$ such that

$$C_1 \leq \sum_{n \in Z} |\hat{\psi}(\alpha^n \omega)|^2 \leq C_2 \quad , \tag{29}$$

then the operator W_d is invertible and has a bounded inverse.

A very important class of discrete wavelet transform was found independently by Y. Meyer [45] and J. Stromberg [54]. They showed that there exist some wavelets $\psi(x) \in \mathbf{L}^2(\mathbf{R})$ such that $\left[\psi_{2^j}(x - 2^{-j} n) \right]_{(n,j) \in Z^2}$ is an orthonormal basis of $\mathbf{L}^2(\mathbf{R})$. These particular wavelets are called *orthogonal wavelets*. Wavelet orthonormal bases can be built for other sequences of scales than $(2^j)_{j \in Z}$ but we will concentrate on dyadic scales which lead to simpler decomposition algorithms. These new orthonormal bases had a striking impact in functional analysis. It was indeed believed that one could not find simple orthonormal bases whose elements have a good localization both in the spatial and frequency domains. Any function can be reconstructed from its decomposition in a wavelet orthonormal basis with the classical expansion formula

$$f(x) = \sum_{j \in Z} \sum_{n \in Z} < f(u) , \psi_{2^j}(u - n 2^{-j}) > \psi_{2^j}(x - n 2^{-j}) \quad . \tag{30}$$

The Haar basis is a well known particular case of wavelet orthonormal basis. The orthogonal wavelet corresponding to the Haar basis is given by

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1/2 \\ -1 & \text{if } 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} . \tag{31}$$

A Haar wavelet is not continuous which is a major inconvenient for many applications. Y. Meyer [45] showed that we can find some orthogonal wavelets $\psi(x)$ which are infinitely continuously differential and have a decay at infinity faster then any power x^{-n} $n > 0$. In paragraph 4.1, we show that the Fourier transform of a large class of orthogonal wavelets can be expressed from the transfer function of a quadrature mirror filter [38]. The decomposition of a function in such a wavelet orthonormal basis can be computed with a quadrature mirror filter bank processing. Fig. 13 gives the graph of a particular orthogonal wavelet and its Fourier transform. This wavelet is a

cubic spline studied independently by Y.Lemarie [33] and G. Battle [3].

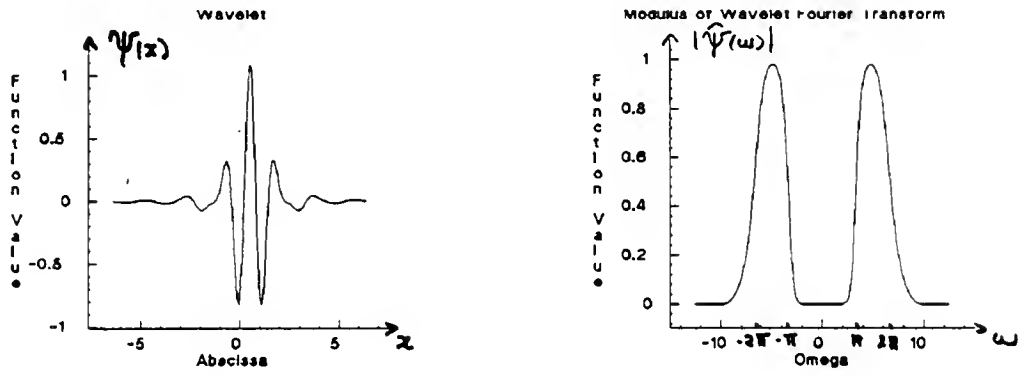


Fig. 13 : *Example of orthogonal wavelet with the modulus of its Fourier transform. It can be interpreted as a band-pass filter. This particular wavelet is a cubic spline.*

An important property of a wavelet transform is to easily characterize the local regularity of a function. In particular, this leads to a simple characterization of the classical functional spaces such as the $\mathbf{L}^p(\mathbf{R})$ spaces, the Sobolev spaces, the Holder spaces ... Let us give an example. One way to measure the regularity of a function is to use the characterization of Holder. A function is inside the functional space of Holder $\mathbf{C}^r$ if and only if $f(x) \in \mathbf{L}^2(\mathbf{R})$ and

$$\forall x_0 \in \mathbf{R} \ , \forall x \in \mathbf{R} \ , \ |f(x) - f(x_0)| \ = o(|x - x_0|^r) \ . \tag{32}$$

The Holder coefficient r gives a measure of the local regularity of a function f(x). The larger r the smoother the function f(x). Y. Meyer and P. Lemarie [34] have shown that if the orthogonal wavelet $\psi(x) \in \mathbf{C}^{r'}$, with $r' > r$, then the functional space $\mathbf{C}^r$ can be characterized by

$$f \in \mathbf{C}^r \iff \exists C > 0 \ , \forall n \ \mathbf{Z} \quad |< f(x) \ , \psi_{2^j}(x - n2^{-j}) >| \ \leq C \ 2^{-j(1/2+r)} \ . \tag{33}$$

The regularity of a function can thus be derived from the decay of the wavelet coefficients when the scale increases. This global characterization can be expressed locally on any neighborhood of a particular point x_1 . Let $\varepsilon > 0$, in the neighborhood $]x_1 - \varepsilon \ , x_1 + \varepsilon[$ of x_1, this property can

be written:

$$\forall x_0 \in]x_1-\varepsilon, x_1+\varepsilon[\ , \forall x \in \mathbf{R} \ , \ |f(x)-f(x_0)| = o(|x-x_0|^r) \ . \tag{34}$$

One can show that the property (34) is valid if and only if there exists a constant $C > 0$ such that for any sequence of integers $(n_j)_{j\in Z}$ which satisfy $\lim_{j\to+\infty} 2^{-j} n_j \in]x_1-\varepsilon, x_1+\varepsilon[$, then

$$|< f(x) , \psi_{2^j}(x-n_j 2^{-j}) >| =< C \ 2^{-j(1/2+r)} \ . \tag{35}$$

The regularity of a function in the neighborhood of a point x_1 thus depend upon the decreasing rate of the wavelet coefficients when the scale increases, in the same neighborhood. Other kind of regularity such as the derivability at any order (in the sense of Sobolev) can be derived similarly [34]. These results show that it is necessary to combine the information at different scales in order to analyze the local properties of a function. In the next paragraph, we describe the extension of the wavelet model to for two-dimensional signals. We come back to orthonormal wavelets in the paragraph 4.1 to explain their relation to the concept of multiresolution in computer vision.

3.5. Wavelet transform in two dimensions

The wavelet transform can be generalized in $\mathbf{R}^n$ but we only consider the two-dimensional case for image processing applications. The model can first be extended without distinguishing any spatial orientation. Let $\Psi(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ be a function such that the integral of its Fourier transform

$$C_\Psi = \int_0^{+\infty} \frac{|\hat{\Psi}(s\,\omega_x, s\,\omega_y)|^2}{s} \, ds \tag{36}$$

is finite and independent from the value of (ω_x, ω_y) . For example, this property is satisfied for a wavelet $\Psi(x,y)$ which is isotropic ($\Psi(x,y) = \rho(\sqrt{x^2+y^2})$) and whose Fourier transform is null in zero ($\hat{\Psi}(0,0) = 0$) . For normalization purposes we suppose that $||\Psi|| = 1$. The function $\Psi(x,y)$ can be interpreted as a band pass filter with no preferential spatial orientation. The

wavelet transform of a function $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ at a scale s , at a point (u,v) , is defined by

$$Wf(s,(u,v)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\, s\, \Psi(s(x-u),s(y-v))\, dx\, dy \quad . \tag{37}$$

Let $\Psi_s(x,y) = s\, \Psi(sx,sy)$ and $\tilde{\Psi}_s(x,y) = \Psi_s(-x,-y)$. The wavelet transform of f(x,y) at the scale s and at a point (u,v) can be rewritten

$$Wf(s,(u,v)) = f * \tilde{\Psi}_s(u,v) \quad . \tag{38}$$

It can be interpreted as a two-dimensional band-pass filtering with no orientation selectivity. The wavelet transform in two dimensions has the same properties as a one-dimensional wavelet transform. It is an isometry from $\mathbf{L}^2(\mathbf{R}^2)$ into $\mathbf{L}^2(\mathbf{R}^+ \times \mathbf{R}^2)$:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{0}^{+\infty} |Wf(s,(u,v))|^2\, ds\, du\, dv = C_\Psi \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |f(x,y)|^2\, dx\, dy \quad . \tag{39}$$

We can reconstruct a function f(x,y) from its wavelet transform with a simple two-dimensional extension of equation (24)

$$f(x,y) = \frac{1}{C_\Psi} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{0}^{+\infty} Wf(s,(u,v))\, \Psi_s(x-u,y-v)\, ds\, du\, dv \quad . \tag{40}$$

In two dimensions, a wavelet transform also satisfies a reproducing kernel equation similar to equation (25) .

For image recognition application, it is often necessary to have a decomposition which differentiates the local orientation of the image features. Let us define N wavelet functions $\Psi^i(x,y)$ $(1 \le i \le N)$ whose Fourier transform $\hat{\Psi}^i(\omega_x,\omega_y)$ satisfy

$$\sum_{i=1}^{N} |\hat{\Psi}^i(\omega_x,\omega_y)|^2 = |\hat{\Psi}(\omega_x,\omega_y)|^2 \quad . \tag{41}$$

Fig. 14 shows an example of decomposition of $\hat{\Phi}(\omega_x,\omega_y)$ into the different functions $\hat{\Psi}^i(\omega_x,\omega_y)$. In the example shown in Fig. 14, the decomposition is symmetrical but this is not a constraint of the model. Each function $\Psi^i(x,y)$ can be viewed as a band pass filter having a particular orientation tuning. The wavelet transform within the orientation i is defined by

$$W^i f(s,(u,v)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\, s\, \Psi^i(s(x-u),s(y-v))\, dx\, dy \quad . \tag{42}$$

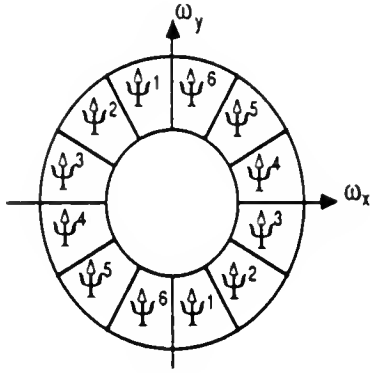


Fig. 14 : *Decomposition in the Fourier domain of the support of* $\hat{\Psi}(\omega_x,\omega_y)$ *in 6 wavelets having different orientation selectivities.*

Let $\Psi_s^i(x,y) = s\ \Psi^i(sx,sy)$ and $\tilde{\Psi}_s^i(x,y) = \Psi_s^i(-x,-y)$. The wavelet transform of f(x,y) at the scale s , at a point (u,v) and within the orientation i can be rewritten:

$$W^i f(s,(u,v)) = f * \tilde{\Psi}_s^i(u,v) \quad . \tag{43}$$

It can thus be interpreted as a convolution of f(x,y) by a band-pass filter having an orientation selectivity. Similarly to equation (37), the wavelet decomposition in several orientations defines an isometry from $\mathbf{L}^2(\mathbf{R}^2)$ into $\mathbf{L}^2(\mathbf{R}^+ \times \mathbf{R}^2)$:

$$\sum_{i=1}^{N} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{0}^{+\infty} |W^i f(s,(u,v))|^2\, ds\, du\, dv = C_\Psi \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |f(x,y)|^2\, dx\, dy \quad . \tag{44}$$

We can also reconstruct a function f(x,y) from its wavelet transform decomposed into several directions:

$$f(x,y) = C_\Psi \sum_{i=1}^{N} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{0}^{+\infty} Wf^i(s,(u,v))\, \Psi_s^i(x-u,y-v)\, ds\, du\, dv \quad .$$

The discretization of a wavelet transform in two dimensions is similar to the discretization in one dimension. We choose a sequence of scales $(\alpha^n)_{n\in Z}$ where α is the elementary dilation step. For each scale α^n , the translation vector (u,v) is uniformly sampled on a two-dimensional grid at a rate proportional to α^n. In the next paragraph, we study the two-dimensional extension of the orthonormal wavelet decomposition and its implementation.

4. Computer vision and multiresolution decompositions

In this paragraph we analyze the multiresolution approach to image interpretation. A multiresolution decomposition is also an image decomposition in frequency channels of constant bandwidth on a logarithmic scale. It provides a different perspective on this kind of transform. We describe the classical pyramidal implementation of multiresolution transforms and show how it relates to a discrete wavelet decomposition.

Multiresolution transform have been thoroughly studied in computer vision since the work of R. Rosenfeld and M. Thurston [50] on multiscale edge detection. At different resolutions, the details of an image generally characterize different types of physical structures. For example, a coarse resolution satellite image of a coast gives a description of only the overall shape of the coast. When the resolution of the image is increased, we are able to successively distinguish the local relief of the region, and if the resolution gets even finer we can recognize the different types of local vegetation. In order to process separately these different structures, researchers in computer vision have tried to extract the difference of information between the approximation of an image at two different resolutions. Given a sequence of increasing resolutions $(r_j)_{j\in Z}$, the details of f(x) at the resolution r_j are defined as the difference of information between the approximation of f(x) at the resolution r_{j+1} and the approximation at the resolution r_j .

A multiresolution representation also provides a simple hierarchical framework for interpreting the image information [30]. In some sense, the details of the image at a coarse resolution provide the "context" of the image whereas the finer details correspond to the particular

"modalities". For example, it is difficult to recognize that a small rectangle inside an image is the window of a house if we did not previously recognize the house "context". It is therefore natural to analyze first the image details at a coarse resolution and then increase the resolution. This is called a coarse to fine processing strategy. At a coarse resolution, the image details are characterized by very few samples. Hence, the coarse information processing can be performed quickly. The finer details are characterized by more samples but the prior information, derived from the context, constrains and thus speeds up the computations. With a coarse to fine strategy, we process the minimum amount of details which are necessary to perform a recognition task. Indeed, if we can recognize an object from a coarse description, we do not need to analyze the finer details. For example, in order to distinguish a car from a house, the coarse details of the image should be enough. Such a strategy is efficient for pattern recognition algorithms. It has already been widely studied for low level image processing such as stereo matching and template matching [21, 24].

4.1. Pyramidal multiresolution decompositions

The approximation of a signal $f(x)$ at a resolution r is defined as an estimate of $f(x)$ derived from r measurements per length unit. These measurements are computed by sampling uniformly at a rate r the function $f(x)$ smoothed by a low-pass filter whose bandwidth is proportional to r. In order to be consistent when the resolution varies, these low-pass filters are derived from a unique function $\theta(x)$ which is dilated by the resolution factor r : $\theta_r = \sqrt{r}\ \theta(rx)$. The set of measurements $A_r f = \left(f * \theta_r(\frac{n}{r}) \right)_{n \in Z}$ is called a discrete approximation of $f(x)$ at the resolution r. In the following we study the approximation of a function on a dyadic sequence of resolution $\left(2^j \right)_{j \in Z}$. The discrete approximation of a function $f(x)$ at the resolution 2^j is thus given by

$$A_{2^j} f = \left(f * \theta_{2^j}(\frac{n}{2^j}) \right)_{n \in Z} . \tag{45}$$

S. Tanimoto and T. Pavlidis [55], P. Burt [4] and J. Crowley [10] have developed efficient pyramidal algorithms to compute the approximation of a function at different resolutions. We first describe these decompositions and then explain P. Burt and J. Crowley algorithm's for computing the details which appear at a resolution 2^{j+1} and do not exist at the resolution 2^j. This simple and elegant algorithm does not not define the details from the difference of information between $A_2^{j+1}f$ and $A_{2^j}f$. It leads to a correlated representation called the Laplacian pyramid. We then review the multiresolution wavelet model which shows that the difference of information between two successive resolutions can be computed by decomposing the signal in a wavelet orthonormal basis.

In pyramidal multiresolution algorithms, the low-pass filter function $\theta(x)$ is chosen such that its Fourier transform can be written

$$\hat{\theta}(\omega) = \prod_{p=1}^{+\infty} U(2^{-p} e^{-i\omega}) \quad , \tag{46}$$

where $U(e^{-i\omega})$ is the transfer function of a low-pass discrete filter $U = \left[u_n \right]_{n \in Z}$. I. Daubechies [11] studied the regularity and decay at infinity of the function $\theta(x)$ depending upon the properties of the filter $U(e^{-i\omega})$. In general we want to have a function $\theta(x)$ which is as smooth as possible and which is well concentrated around 0 in the spatial domain. I. Daubechies proved that an optimal choice of the impulse response $\left[u_n \right]_{n \in Z}$ is given by the coefficients of a binomial polynomial.

Let us suppose that we have already computed the discrete approximation of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ at the resolution 2^j : $A_{2^j}f = \left[f * \theta_{2^j}(\frac{n}{2^j}) \right]_{n \in Z}$. One can show [4, 11, 38] that the discrete approximation of f(x) at a resolution 2^{j-1} is calculated by filtering $A_{2^j}f$ with the discrete low-pass filter $U = \left[u_n \right]_{n \in Z}$ and keeping every other sample of the convolution product. Let $\Lambda = \left[\lambda_n \right]_{n \in Z}$ be such that

$$\Lambda = A_{2^j} f \; * \; U \;, \tag{47}$$

$$\text{then} \quad A_{2^{j-1}} f \; = \left(\lambda_{2n} \right)_{n \in Z} \;. \tag{48}$$

A measuring device provides the approximation of an input signal at a finite resolution. Let us suppose for normalization purposes that this resolution is equal to one. The approximation of this signal at any resolution 2^{-J} , $J > 0$, can be computed by iterating on equation (47) for j varying between 0 and J+1 . This pyramidal algorithm is illustrated in Fig. 15(a). The set of discrete approximations $\left(A_{2^j} f \right)_{0 \geq j \geq -J}$ was called a Gaussian pyramid by P. Burt [4].

We now describe the algorithm of P. Burt [4] and J. Crowley [10] in order to extract the details of f(x) which appear in $A_{2^j} f$ but not in $A_{2^{j-1}} f$. The discrete approximation $A_{2^j} f$ has twice more samples than $A_{2^{j-1}} f$ so we first expand $A_{2^{j-1}} f$ by a factor two. This is performed with a classical interpolation procedure [9]. We put a zero between each sample of $A_{2^{j-1}} f$ and filter the resulting signal with a low-pass filter. In this algorithm, the low-pass filter is the filter U defined previously. Let $A_{2^{j-1}} f_e$ be the expanded discrete signal. The details $D_{2^j} f$ at the resolution 2^j are then computed by subtracting $A_{2^{j-1}} f_e$ from $A_{2^j} f$:

$$D_{2^j} f \; = A_{2^{j-1}} f \; - A_{2^j} f_e \;. \tag{49}$$

This algorithm decomposes a discrete approximation $A_1 f$ at a resolution 1 into an approximation $A_{2^{-J}} f$ at a coarse resolution 2^{-J} and the successive detail signals $\left(D_{2^j} f \right)_{0 < j \leq -J}$. If the signal $A_1 f$ has N non-zero samples, each detail signal $D_{2^j} f$ has $2^{j+1} N$ samples whereas the coarse signal $A_{2^{-J}} f$ has $2^{-j} N$ samples. Hence, the total number of samples of this representation is approximatively $2N$. Such a representation is often called a Laplacian pyramid [4].

The original signal can easily be reconstructed from such a decomposition. At each resolution we compute $A_{2^{j-1}} f$ by expanding $A_{2^j} f$ by a factor two and adding the details $D_{2^j} f$. By repeating this algorithm when j is varying between -J and 0 we reconstruct $A_1 f$. The reconstruction algorithm is illustrated by a block diagram in Fig. 15(b).

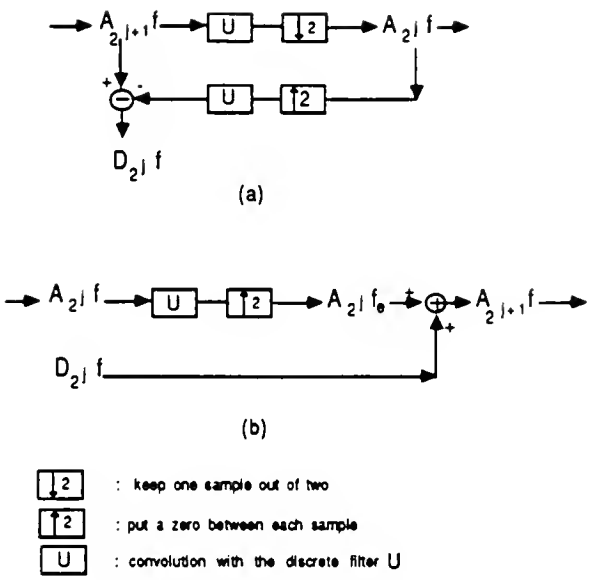


Fig. 15 : *(a): Decomposition of $A_{2^{j+1}}f$ into $A_{2^j}f$ and $D_{2^j}f$ when computing a Laplacian pyramid. (b): Reconstruction of $A_{2^{j+1}}f$ from $A_{2^j}f$ and $D_{2^j}f$ when reconstructing the original signal from a Laplacian pyramid.*

In two dimensions, the discrete approximation of a signal $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ at the resolution 2^j is similarly defined by

$$A_{2^j}f = \left[f * \Theta_{2^j}(2^{-j}n, 2^{-j}m) \right]_{(n,m)\in Z^2} ,$$

where $\Theta(x,y)$ is a two-dimensional low-pass filter and $\Theta_{2^j}(x,y) = 2^j\,\Theta(2^j x, 2^j y)$. For image processing, the pyramidal algorithm is extended with separable convolutions along the rows and columns of the image [4]. The low-pass filter $\Theta(x,y)$ is chosen such that its Fourier transform can be written :

$$\hat{\Theta}(\omega_x,\omega_y) = \prod_{p=1}^{+\infty} U(2^{-p}e^{-i\omega_x})\,U(2^{-p}e^{-i\omega_y}) \quad . \tag{50}$$

Let us suppose that the video camera provides an image approximated at the resolution 1 : $A_1 f = \left[f * \Theta(n,m) \right]_{(n,m)\in Z^2}$. With a separable extension of the algorithm described in

equations (47) and (48), we can compute the approximation of an image at any resolutions 2^j, j < 0 . Fig. 16 shows an image approximated at the resolution 2^j for $0 \geq j \geq -3$ (Gaussian pyramid) . The detail signals $\left[D_{2^j} f\right]_{0<j\leq -3}$ can also be computed with a straight forward extension of the one-dimensional algorithm. Fig. 17 shows the discrete signals (Laplacian pyramid) corresponding to the image given in Fig. 16. If the original image has N^2 pixels, each detail image $D_{2^j} f$ has $2^{j+1}N^2$ pixels and $A^{-J} f$ has $2^{-J}N^2$ pixels. Hence, the total number of pixels of this representation is approximatively $\frac{4}{3}N^2$. This pyramidal algorithm does not introduce any orientation selectivity in the decomposition process.

Fig. 16 : *Gaussian pyramid. The image is approximated at the resolutions* 1 , $\frac{1}{2}$, $\frac{1}{4}$ *and* $\frac{1}{8}$. *When the resolution decreases with the loose the details and the image is characterized by less pixels.*

In a Laplacian pyramid, the signals $D_{2^j} f$ do not correspond to the difference of information between $A_{2^{j-1}} f$ and $A_{2^j} f$. If this was the case, the total number of pixels representing the signal would be the same as in the original signal. We saw that the number of samples representing the signal is increased by a factor 2 in one dimension and by a factor $\frac{4}{3}$ in two dimensions. It is due to the correlation between the detail signals $D_{2^j} f$ at different resolutions. This

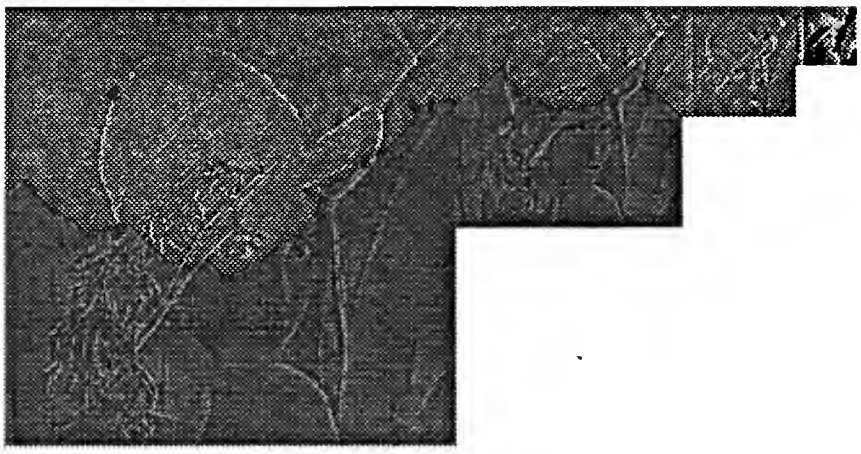


Fig. 17 : *Laplacian pyramid. This figure shows the detail images at the resolution* $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ *and the coarse image approximated at the resolution* $\frac{1}{8}$.

correlation can be understood and suppressed with the multiresolution wavelet model described in [39,38]. It is indeed possible to extract exactly the difference of information between $A_{2^{j+1}}f$ and $A_{2^j}f$ by decomposing the signal in a wavelet orthonormal basis.

Let us first explain the multiresolution wavelet model in one dimension. We saw in equation (45) that the discrete approximation of a function f(x) at the resolution 2^j is defined by $A_{2^j}f = \left[f * \theta_{2^j}(2^{-j}n) \right]_{n \in \mathbf{Z}}$. Let us denote $\tilde{\theta}_{2^j}(x) = \theta_{2^j}(-x)$. Each convolution product in a point can be rewritten as an inner product in $\mathbf{L}^2(\mathbf{R})$:

$$A_{2^j}f = \left[< f(x), \tilde{\theta}_{2^j}(x-2^{-j}n) > \right]_{n \in \mathbf{Z}} . \tag{51}$$

From such a set of inner products, the best estimate of f(x) which can be computed is the orthogonal projection of f(x) on the vector space $\mathbf{V}_{2^j}$ generated by the family of function $\left[\tilde{\theta}_{2^j}(x - 2^{-j}n) \right]_{n \in \mathbf{Z}}$. Let us call the *continuous approximation* at the resolution 2^j this

orthogonal projection. The vector spaces $\mathbf{V}_{2^j}$ can be viewed as all the possible approximations of $\mathbf{L}^2(\mathbf{R})$ functions at the resolution 2^j. The sequence of vector spaces $\left(\mathbf{V}_{2^j}\right)_{j\in\mathbf{Z}}$ is called a *multiresolution approximation of* $\mathbf{L}^2(\mathbf{R})$. The properties of the vector spaces $\mathbf{V}_{2^j}$ are further studied in [39, 38]. For any function $f(x) \in \mathbf{L}^2(\mathbf{R})$, the best estimate of f(x) at the resolution 2^j is given by the orthogonal projection of f(x) on $\mathbf{V}_{2^j}$. In order to compute this estimate, we need to have an orthonormal basis of $\mathbf{V}_{2^j}$. One can show [39] that we can build such an orthonormal basis by dilating and translating a particular function $\phi(x)$ called a *scaling function*. Let $\phi_{2^j}(x) = \sqrt{2^j}\,\phi(2^j x)$, the family of functions $\left(\phi_{2^j}(x-2^{-j}n)\right)_{n\in\mathbf{Z}}$ is an orthonormal basis of $\mathbf{V}_{2^j}$. The Fourier transform of $\phi(x)$ can be written

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} H(2^{-p} e^{-i\omega}) \quad , \tag{52}$$

where $H(e^{-i\omega})$ is the transfer function of a discrete filter. One can show [39] that $H(e^{-i\omega})$ satisfies the condition

$$|H(e^{-i\omega})|^2 + |H(e^{i\omega})|^2 = 1 \quad . \tag{53}$$

The discrete filters $H = \left(h_n\right)_{n\in\mathbf{Z}}$ whose transfer function satisfy equation (53) are called quadrature mirror filters [15].

The orthogonal projection of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ on $\mathbf{V}_{2^j}$ can now be computed by decomposing $f(x)$ on the orthonormal basis $\left(\phi_{2^j}(x-2^{-j}n)\right)_{n\in\mathbf{Z}}$. Let $\mathbf{P}_{\mathbf{V}_{2^j}}$ be the orthogonal projection operator on $\mathbf{V}_{2^j}$.

$$\mathbf{P}_{\mathbf{V}_{2^j}}(f)(x) = \sum_{n\in\mathbf{Z}} \langle f(u), \phi_{2^j}(u-2^{-j}n)\rangle \;\; \phi_{2^j}(x-2^{-j}n) \quad . \tag{54}$$

Let us denote $\tilde{\phi}(x) = \phi(-x)$. Since $\phi(x)$ is a low-pass filter, we can redefine the discrete approximation $A_{2^j}f$ with the function $\tilde{\phi}(x)$ instead of $\theta(x)$:

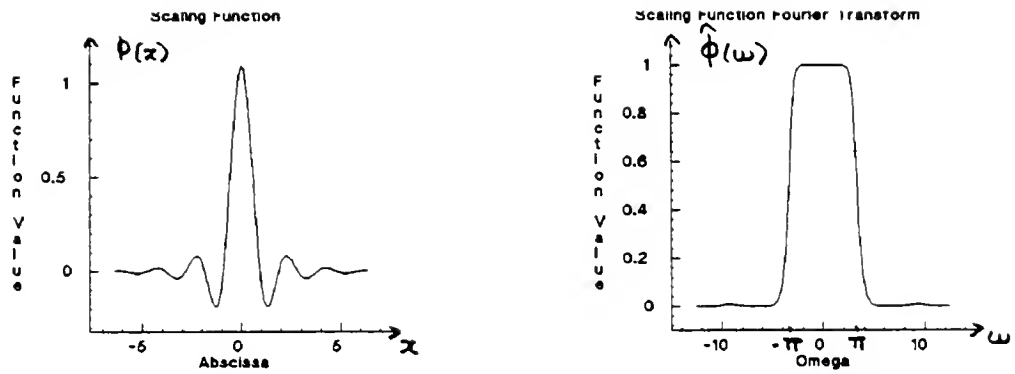


Fig. 18 : *Example of scaling function with its Fourier transform. A scaling function is a low-pass filter. The computation of this particular function is described in [38]. The corresponding orthogonal wavelet is shown in Fig. 13.*

$$A_{2^j} f = \left[f * \bar{\phi}_{2^j}(2^{-j} n) \right]_{n \in \mathbf{Z}} = \left[\langle f(x), \phi_{2^j}(x - 2^{-j} n) \rangle \right]_{n \in \mathbf{Z}} . \tag{55}$$

The best estimate of f(x) can easily be derived from this discrete approximation by using equation (54). Let $\bar{H}$ be the discrete filter whose impulse response is $\left[h_{-n} \right]_{n \in \mathbf{Z}}$. From equations (52) and (55), one can show [38] that the discrete approximations $A_{2^j} f$ is computed with the same pyramidal algorithm described in equations (47) and (48) using the discrete filter $\bar{H}$ instead of U.

Let us now explain how to extract exactly the difference of information between the approximations of a function at the resolutions 2^j and 2^{j+1}. The approximations of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ at the resolutions 2^j and 2^{j+1} are respectively given by the orthogonal projection of f(x) on the vector spaces $\mathbf{V}_{2^j}$ and $\mathbf{V}_{2^{j+1}}$. The approximation at the resolution 2^{j+1} is a better estimate of f(x) than the approximation at the resolution 2^j. Hence, the vector spaces $\mathbf{V}_{2^j}$ and $\mathbf{V}_{2^{j+1}}$ verify

$$\mathbf{V}_{2^j} \subset \mathbf{V}_{2^{j+1}} . \tag{56}$$

By applying the projection theorem, we derive that the difference of information between the two approximations is equal to the orthogonal projection of f(x) on the orthogonal complement of $\mathbf{V}_{2^j}$ in $\mathbf{V}_{2^{j+1}}$. Let $\mathbf{O}_{2^j}$ be this orthogonal complement. To compute the orthogonal projection of a function f(x) on $\mathbf{O}_{2^j}$, we need to find an orthonormal basis of $\mathbf{O}_{2^j}$. One can show [39] that such an orthonormal basis can be built by dilating and translating a particular wavelet $\psi(x)$. Let $\psi_{2^j}(x) = \sqrt{2^j}\,\psi(2^j x)$, the family of functions $\left(\psi_{2^j}(x-2^{-j}n)\right)_{n\in\mathbf{Z}}$ is an orthonormal basis of $\mathbf{O}_{2^j}$. The Fourier transform of $\psi(x)$ is given by

$$\hat{\psi}(2\omega) = G(e^{-i\omega})\,\hat{\phi}(\omega) \qquad \text{with} \qquad G(e^{-i\omega}) = e^{-i\omega}\,\overline{H(-e^{-i\omega})}\,. \tag{57}$$

$G(e^{-i\omega})$ is the transfer function of a discrete filter $G = \left(g_n\right)_{n\in\mathbf{Z}}$. The filters G and H make a pair of quadrature mirror filters [53].

When the resolution 2^j varies between 0 and $+\infty$, the family of functions $\left(\psi_{2^j}(x-2^{-j}n)\right)_{(n,j)\in\mathbf{Z}^2}$ constitutes a wavelet orthonormal of $\mathbf{L}^2(\mathbf{R})$ [39]. This shows that the multiresolution concept and quadrature mirror filters are directly related to wavelet orthonormal bases. Equations (52) and (53) provide a simple way to compute new orthogonal wavelets. Indeed, quadrature mirror filters are not difficult to synthesize.

Let $\mathbf{P}_{\mathbf{O}_{2^j}} f(x)$ be the orthonormal projection of a function $f(x) \in \mathbf{L}^2(\mathbf{R})$ on the vector space $\mathbf{O}_{2^j}$. $\mathbf{P}_{\mathbf{O}_{2^j}} f(x)$ gives the difference of information between the approximations of f(x) at the resolutions 2^j and 2^{j+1}. It can be computed by expanding f(x) in the orthonormal basis of $\mathbf{O}_{2^j}$:

$$\mathbf{P}_{\mathbf{O}_{2^j}} f(x) = \sum_{n\in\mathbf{Z}} \langle f(u)\,,\,\psi_{2^j}(u-2^{-j}n)\rangle \;\; \psi_{2^j}(x-2^{-j}n)\,. \tag{58}$$

This difference of information is thus characterized by the set of inner products

$$D_{2^j} f = \left(\langle f(x)\,,\,\psi_{2^j}(x-2^{-j}n)\rangle\right)_{n\in\mathbf{Z}}\,. \tag{59}$$

Hence, at each scale 2^j, the decomposition of a function f(x) in a wavelet orthonormal basis is

equal to the difference of information between the approximation of f(x) at the resolutions 2^{j+1} and 2^j . Let $\bar{G}$ be the filter whose impulse response is given by $\bar{G} = \left[g_{-n} \right]_{n \in \mathbf{Z}}$. From equations (55) , (57) and (59), one can derive that $D_{2^j} f$ is computed by filtering $A_{2^j} f$ with $\bar{G}$ and keeping every other sample of the convolution product [38]. This algorithm is illustrated by the block diagram shown in Fig. 19(a); it is essentially similar to a quadrature mirror filter bank decomposition [15].

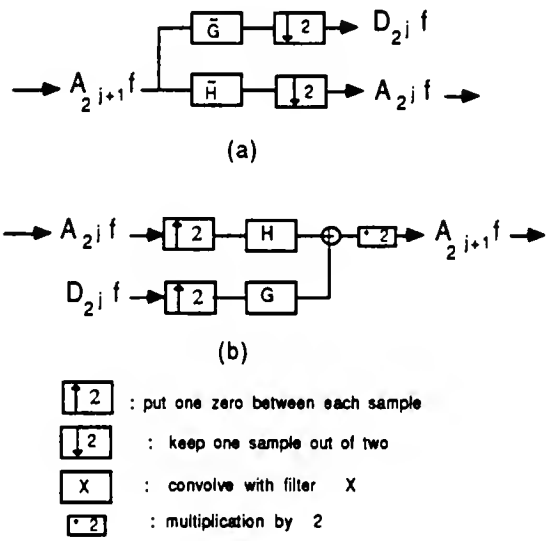


Fig. 19 : *(a): Decomposition of $A_{2^{j+1}} f$ into $A_{2^j} f$ and $D_{2^j} f$ when computing an orthogonal wavelet representation. The filters H and G make a pair of quadrature mirror filters. (b): Reconstruction of $A_{2^{j+1}} f$ from $A_{2^j} f$ and $D_{2^j} f$ when reconstructing the original signal from an orthogonal wavelet representation.*

Let us now describe a simple two-dimensional extension of the one-dimensional multiresolution wavelet model. We saw that a separable multiresolution representation is computed by filtering the signal with a low-pass filter $\Theta(x,y) = \theta(x)\,\theta(y)$. The discrete approximation of a function $f(x,y) \in \mathbf{L}^2(\mathbf{R}^2)$ at the resolution 2^j is defined by

$$A_{2^j} f = \left[\langle f(x,y) , \Theta_{2^j}(x - 2^{-j} n , y - 2^{-j} m) \rangle \right]_{(n,m) \in \mathbf{Z}^2} . \qquad (60)$$

Let $\tilde{\Theta}(x,y) = \Theta(-x,-y)$. With a straight-forward extension of the one-dimensional model, we can show that the best estimate of f(x,y) at the resolution 2^j is given by its orthogonal projection on the vector space $\mathbf{V}_{2^j}$ generated by the family of functions

$$\left\{ \Theta_{2^j}(x-2^{-j}n\ ,\ y-2^{-j}m) \right\}_{(n,m)\in \mathbf{Z}^2} \quad . \tag{61}$$

The sequence of vector spaces $\left\{ \mathbf{V}_{2^j} \right\}_{j\in \mathbf{Z}}$ is called a multiresolution approximation of $\mathbf{L}^2(\mathbf{R}^2)$. Similarly to the one-dimensional model, the difference of information between the approximation of a signal f(x,y) at the resolutions 2^j and 2^{j+1} is equal to the orthogonal projection of f(x,y) on the orthogonal complement $\mathbf{O}_{2^j}$ of $\mathbf{V}_{2^j}$ in $\mathbf{V}_{j+1}$. One can show [45] that we can build an orthonormal basis of $\mathbf{O}_{2^j}$ by scaling and translating three wavelets : $\Psi^1(x,y)$, $\Psi^2(x,y)$ and $\Psi^3(x,y)$. These wavelets are given by :

$$\Psi^1(x,y) = \phi(x)\,\psi(y)\ ,\ \Psi^2(x,y) = \psi(x)\,\phi(y)\ ,\ \Psi^3(x,y) = \psi(x)\,\psi(y)\ , \tag{64}$$

where $\psi(x)$ is the one-dimensional orthogonal wavelet defined by equation (57). Let $\Psi^k_{2^j}(x,y) = 2^j\ \Psi^k(2^j x, 2^j y)$ for $1 \le k \le 3$. The family of functions

$$\left\{ \begin{array}{l} 2^{-j}\ \Psi^1_{2^j}(x-2^{-j}n\ ,\ y-2^{-j}m) \\ 2^{-j}\ \Psi^2_{2^j}(x-2^{-j}n\ ,\ y-2^{-j}m) \\ 2^{-j}\ \Psi^3_{2^j}(x-2^{-j}n\ ,\ y-2^{-j}m) \end{array} \right\}_{(n,m)\in \mathbf{Z}^2} \tag{57}$$

is an orthonormal basis of $\mathbf{O}_{2^j}$. When the resolution 2^j varies between 0 and $+\infty$, the family of functions

$$\left\{ \begin{array}{l} 2^{-j}\ \Psi^1_{2^j}(x-2^{-j}n\ ,\ y-2^{-j}m) \\ 2^{-j}\ \Psi^2_{2^j}(x-2^{-j}n\ ,\ y-2^{-j}m) \\ 2^{-j}\ \Psi^3_{2^j}(x-2^{-j}n\ ,\ y-2^{-j}m) \end{array} \right\}_{(n,m,j)\in \mathbf{Z}^3} \tag{66}$$

is a wavelet orthonormal basis of $\mathbf{L}^2(\mathbf{R}^2)$. Fig. 20 shows approximatively the frequency support of the three wavelets $\Psi^1(x,y)$, $\Psi^2(x,y)$, $\Psi^3(x,y)$. Each filter $\Psi^i(x,y)$ is a band-pass filter having a specific orientation selectivity.

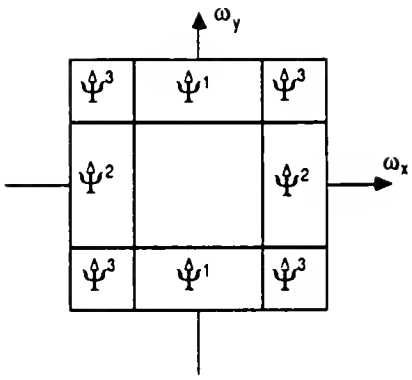


Fig. 20 : *Approximative repartition of the frequency support of* $\hat{\Psi}^1(\omega_x,\omega_y)$, $\hat{\Psi}^2(\omega_x,\omega_y)$ *and* $\hat{\Psi}^3(\omega_x,\omega_y)$.

In two dimensions, the difference of information between the approximations $A_{2^{j+1}}f$ and $A_{2^j}f$ is therefore characterized by the sequences of inner products :

$$D^1_{2^j}f = \left[< f(x,y) , \Psi^1_{2^j}(x-2^{-j}n , y-2^{-j}m) > \right]_{(n,m)\in Z^2} , \tag{67}$$

$$D^2_{2^j}f = \left[< f(x,y) , \Psi^2_{2^j}(x-2^{-j}n , y-2^{-j}m) > \right]_{(n,m)\in Z^2} , \tag{68}$$

$$D^3_{2^j}f = \left[< f(x,y) , \Psi^3_{2^j}(x-2^{-j}n , y-2^{-j}m) > \right]_{(n,m)\in Z^2} . \tag{69}$$

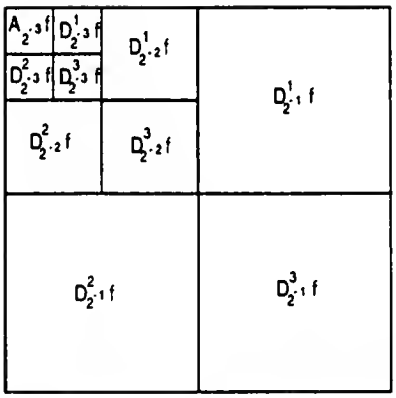
Each of these sequences of inner products can be considered as an image. $D^1_{2^j}f$ gives the vertical higher frequencies (horizontal edges), $D^2_{2^j}f$ the horizontal higher frequencies (vertical edges) and $D^3_{2^j}f$ the higher frequencies in both directions (corners) (see Fig. 21). Let us suppose that we have initially an image $A_1 f$ measured at the resolution 1. For any $J > 0$, this discrete image can be decomposed between the resolutions 1 and 2^{-J} , and completely represented by the $3J + 1$ discrete images :

$$\left[A_{2^{-J}}f , (D^1_{2^j}f)_{-J\le j\le -1} , (D^2_{2^j}f)_{-J\le j\le -1} , (D^3_{2^j}f)_{-J\le j\le -1} \right] . \tag{70}$$

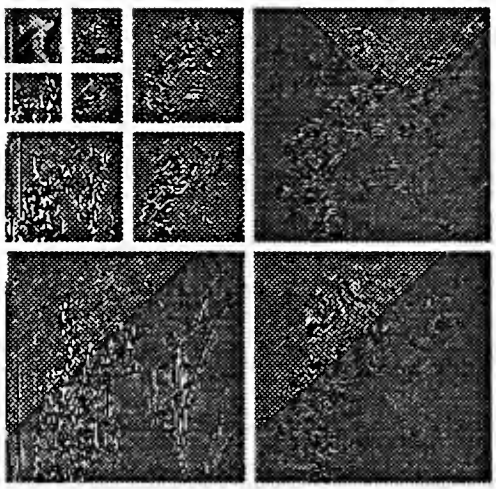
This set of images is called an *orthogonal wavelet representation* in two dimensions [38]. The

image $A_{2^{-j}}f$ is a coarse approximation and the images $D^k_{2^j}f$ give the image details for different orientations and resolutions. If the original image has N^2 pixels, each image $A_{2^j}f$, $D^1_{2^j}f$, $D^2_{2^j}f$, $D^3_{2^j}f$ has $2^j \cdot N^2$ pixels (j < 0). The total number of pixels in this new representation is equal to the number of pixels of the original image, so we do not increase the volume of data. This is due to the orthogonality of the representation.

A wavelet representation can be computed with a separable extension of the algorithm illustrated in Fig. 19(a) [38]. This extension corresponds to a separable quadrature mirror filter decomposition as described by Woods [61]. Fig. 20(b) gives the wavelet representation of the lady image. From this representation we can reconstruct the original image with two-dimensional separable extension of the algorithm illustrated in 19(b) [38]. Fig. 20(c) is the reconstructed image from the wavelet representation shown in Fig. 20(b). The reconstruction is numerically stable. It enables us to use this type of representation for image coding. A more general non-separable extension of the wavelet model was studied by Y. Meyer [43]. Such extensions are however more difficult to implement and are computationally more expensive.



(a)



(b)

(c)

Fig. 21 : *(a): Labelling of the detail images shown in the wavelet representation. (b): Wavelet representation of the lady image. (c): Reconstruction of the original image from the orthogonal wavelet representation.*

4.2. Applications of multiresolution transforms

The wavelet model gives a precise understanding of the concept of multiresolution by introducing the sequence of vector spaces $\left\{ V_j \right\}_{j \in Z}$. A non-correlated multiresolution representation can be built by decomposing the signal in a wavelet orthonormal basis. A difficult problem when using a multiresolution representation for analyzing a scene is to relate the details appearing at different resolutions. Many ad-hoc technics have been developed for this purpose. We saw in paragraph 3.4 , that the local regularity of a function is provided by the decreasing rate of the wavelet coefficients when the resolution increases. These theorems give a first approach for comparing the value of the decomposition at different resolutions.

Multiband image decompositions are also well adapted for coding images because it is possible to match the human visual system sensitivity and take advantage of the intrinsic statistical properties of images. The contrast sensitivity function (Fig. 3) shows that the sensitivity of human vision depends upon the frequency of the stimulus. We want to quantize each frequency band with the minimum number of bits and at the same time try to reconstruct the best possible image for human visual perception. For this purpose, we adapt the quantization noise to the human sensitivity along each frequency band. The more sensitive the human system, the less quantization noise is introduced. This enables us to introduce a minimum amount of perceivable distortion in the reconstructed image. A. Watson made some particularly detailed psychophysical experiments to test this type of approach for image coding [58].

The statistical properties of images give another reason for using multiband decompositions in image coding. It is well known that the intensity of images is locally correlated. Predictive codings have been particularly successful to compress the number of bits to code an image. Recently, Markov random field models have been developed to modelize the local statistical properties of the image intensity [18]. The wavelet coefficients give a measure of the image local contrast at different scales. Since the image intensity is locally correlated, in general these local

contrast have a relatively small amplitude [38]. We can take advantage of this property for coding the wavelet coefficients on fewer bits without introducing a noticeable distortion. As explained in the previous paragraph, a wavelet orthogonal representation can also be interpreted as a a decomposition in a quadrature mirror filter bank. Several studies in image processing have already shown the efficiency of these filter banks for data compression [1, 61].

In order to use a multiresolution representation for pattern recognition applications, we must be able to build models of patterns within the multiresolution representation. The patterns might be located anywhere in the image. Hence, the models must be independent from the pattern location. When a pattern is translated, its model should only be translated but not modified. Let us show that a multiresolution representation does not verify this translating property. To simplify the explanation, we consider the particular case of a one-dimensional orthogonal wavelet decomposition. At the resolution 2^j, the details of a signal $f(x) \in \mathbf{L}^2(\mathbf{R})$ are defined by

$$D_{2^j} f = \left[< f(x) , \psi_{2^j}(x - 2^{-j} n) > \right]_{n \in \mathbf{Z}} .$$

$D_{2^j} f$ can be expressed as a uniform sampling of the wavelet transform at the scale 2^j :

$$D_{2^j} f = \left[W f(2^j, 2^{-j} n) \right]_{n \in \mathbf{Z}} .$$

Let $g(x) = f(x - \tau)$ be a translation of f(x) by τ. Since a wavelet transform can be written as a convolution product (equation (22)), it is shift invariant :

$$W g(2^j, u) = W f(2^j, u - \tau) .$$

However, the sampling of $W g(2^j, u)$ does not correspond to a translation of the sampling of $W f(2^j, u)$ unless $\tau = k 2^j$, $k \in \mathbf{Z}$ (see Fig. 22) . This distortion through translation implies that the wavelet coefficients of a pattern at the resolution 2^j depend upon the position of the pattern modulo 2^{-j} . This property is inherent to the notion of resolution. Indeed, at the resolution 2^j , we cannot measure anything smaller than 2^{-j} so we cannot represent a displacement smaller

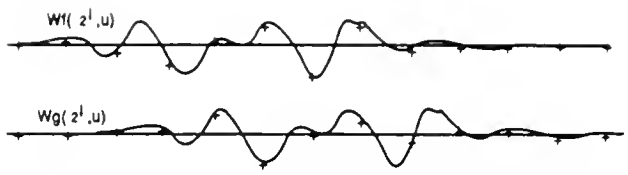


Fig. 22 : *This drawing shows that the sampling of a wavelet transform (given by the crosses) can be very different after translating the signal. The sampling does not translate if the translation is not proportional to the sampling rate (adapted from [40]).*

than 2^{-j}. One can find the same problem in all the pyramidal multiresolution representations and any uniform sampling of a wavelet transform.

A first solution to this translation problem is to sample the wavelet transform $Wf(2j,u)$ at a rate higher than 2^j. The samples will then translate approximatively when the signal translates. However, this solution increases considerably the redundancy of the representation and the translation is still not perfect. This technique is often adopted for pattern recognition algorithms based on pyramid decompositions. A second solution consists in defining a representation based on an adaptive sampling of the functions $Wf(2^j,u)$, which translates when the signal translates.

5. Zero-crossings of multifrequency channels

In the previous paragraphs we studied the properties of the decomposition of a function decomposition in multifrequency channels of constant size on a logarithmic scale. We saw that such a decomposition can be interpreted as a wavelet transform. We then described the properties and applications of a discrete wavelet transform built from a uniform sampling of the continuous wavelet transform. However, we showed that such a discretization is difficult to use for pattern recognition applications because it is not invariant through translation. In this paragraph, we review the characterization of a signal from the zero-crossings of a wavelet transform. Indeed, such a characterization defines a discrete representation which translates when the signal

translates.

If a function f(x) is translated, for each scale s , the function $Wf(s,u)$ is translated along the parameter u . Hence, the zero-crossings of $Wf(s,u)$ are translated as well. If $\psi(x)$ is equal to the second derivative of a low-pass filter $\xi(x)$, any zero-crossing of $Wf(s,u)$ can be interpreted as a point of abrupt change in the function f(x) smoothed by $\xi_s(x) = \sqrt{s}\ \xi(sx)$. Indeed, if $\psi(x) = \xi''(x)$:

$$Wf(s,u) = f * \bar{\psi}_s(u) = s^2 (f * \xi_s)''(u) .$$

Hence, a zero-crossing of $Wf(s,u)$ corresponds to an inflection point of the function f(x) smoothed by $\xi_s(x)$. Fig. 23 illustrates this on a straight edge.

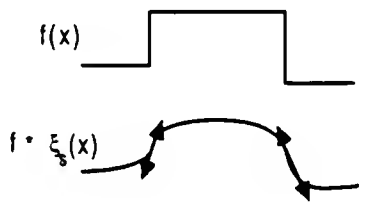


Fig. 23 : *The zero-crossings of a wavelet transform provide the location of the inflection points (edges) of* $f * \xi_s(x)$ *(adapted from [40]).*

Let us now study the completeness and stability of such a representation. We know that a wavelet transform $Wf(s,u)$ defines a stable and complete representation of f(x) . It is therefore equivalent to study the reconstruction of $Wf(s,u)$ from its own zero-crossings. If the function $Wf(s,u)$ was a priori any function of $\mathbf{L}^2(\mathbf{R}^+ \times \mathbf{R})$, it is clear that such a reconstruction would not be possible. Indeed, for a given set of zero-crossings, there is an infinite number of functions in $\mathbf{L}^2(\mathbf{R}^+ \times \mathbf{R})$ whose zero-crossings correspond to this set. However, we saw that a wavelet transform $Wf(s,u)$ is not any function of $\mathbf{L}^2(\mathbf{R}^+ \times \mathbf{R})$. It verifies the constraint of the reproducing kernel equation (25). We must therefore study whether the constraint of the reproducing kernel plus the information on the zero-crossing positions is enought to have a stable

characterization of $Wf(s,u)$.

An interesting particular case of wavelet transform consists in choosing a wavelet equal to the Laplacian of a Gaussian. Since a Gaussian is a low-pass filter, the zero-crossings of such a wavelet transform can be interpreted as signal edges. Marr and Hildreth [41] have particularly studied these zero-crossings for edge detection. In this particular case the intrinsic redundancy of the wavelet transform $Wf(s,u)$ can be expressed with the differential equation of heat diffusion [30]. By applying the maximum principles to the solutions of the heat differential equation, R.Hummel [26] proved that a function f(x) is indeed characterized by the zero-crossings of $Wf(s,u)$. However, R. Hummel showed also that this characterization is not stable. To a slight perturbation of the zero-crossings may correspond a substantial perturbation of the corresponding function f(x). Reconstruction algorithms have been developed on images by J. Sanz and T. Huang [52] as well as Y. Zeevi and D. Rotem [62]. These reconstruction algorithms are iteratives. They were not able to perfectly reconstruct the image in both cases. R. Hummel and R. Moniot [27] tried to stabilize the zero-crossing representation by also recording the value of the gradients of $Wf(s,u)$ along each zero-crossing. By adding the gradient information, they have shown experimentally that one can then compute a stable reconstruction of f(x) from the zero-crossings of $Wf(s,u)$. In this algorithm, the position of the zero-crossings and the value of the gradients are kept along a uniform sequence of scales : $\left(j\alpha\right)_{j\in Z}$ with $\alpha > 0$. Such a sequence is much more dense than the dyadic sequences $\left(2^j\right)_{j\in Z}$ used when we discretized uniformly the wavelet transform.

Another way to stabilize a zero-crossing representation is to record the energy of $Wf(s,u)$ between two consecutive zero-crossings appearing at the same scale [40]. This energy preserves an $L^2(R)$ structure to the zero-crossing representation. In particular, we can then define an $L^2(R)$ distance for pattern recognition applications. By keeping the position of the zero-crossings of $Wf(s,u)$ and the local energies only along a dyadic sequence of scales $\left(2^j\right)_{j\in Z}$, we showed that the original can be reconstructed with very few iterations [40]. The

reconstruction uses the reproducing kernel equation which is valid for any type of wavelet transform.

Representations based on zero-crossing of multifrequency channels are still not well understood. They are built with a non linear transform which is difficult to modelize. However, they have very good potentials for pattern characterizations. Indeed, they characterize the position of the signal edges and are translational invariant.

6. Conclusion

In this paper, we reviewed from different point of views the application of multifrequency decompositions to image processing. We covered some psychophysical and physiological data showing that such a decomposition seems to be implemented in the human visual cortex. We then described the mathematical properties of these decompositions. We first reviewed the properties of a window Fourier transform and explained why this decomposition is not convenient for analyzing signals such as images. We then introduced the wavelet transform and described its most important properties. Although the goal of this paper was not to build any psychophysiological model of the human visual system, it would be interesting to further investigate the relevance of the wavelet model to some low-level processings in the visual cortex.

In computer vision, multifrequency channel decompositions are interpreted through the concept of multiresolution. We described the classical pyramidal multiresolution algorithms and the wavelet approach to multiresolution decompositions. This model shows that the difference of information between the approximation of a function at two different resolution is computed by decomposing this function in a wavelet orthonormal basis. We also explained the relation between orthonormal wavelets and quadrature mirror filters. We can compute the decomposition of a function in a wavelet orthonormal basis with a quadrature mirror filter bank. A third motivation for using multiband decomposition is due to the intrinsic statistical properties of images. Images have a relatively simple decomposition into frequency subbands. These bands can be

coded on fewer bits with no visible distortions.

A uniform sampling of each multifrequency channel defines a representation which is not translational invariant. It is then difficult to build pattern recognition algorithms from such decompositions. In the last paragraph, we reviewed the zero-crossing sampling of multiband decompositions. This adaptive sampling is translational invariant but its properties are much more difficult to analyze. We described some previous results, and the wavelet formalization of this problem through the reproducing kernel equation.

Acknowledgments

I would like to thank particularly Ruzena Bajcsy for her advice on the writing of this paper and Nicolas Treil for helping me drawing the figures.

References

1. Adelson, E. and Simoncelli, E., "Orthogonal pyramid transform for image coding," *Proc. SPIE, Visual Commun. and Image Proc.*, 1987.

2. Andrew, B. and Pollen, D., "Relationship between spatial frequency selectivity and receptive field profile of simple cells," *J. of Physiology*, vol. 287, pp. 163-176, 1979.

3. Battle, G., "A block spin construction of ondelettes, Part 1 : Lemarie functions," *Comm. Math. Phys.* , vol. 110, pp. 601-615, 1987.

4. Burt, P. J. and Adelson, E. H., "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communications, vol 31, pp 532-540*, April 1983.

5. Campbell, F., Carpenter, R., and Levinson, J., "Visibility of aperiodic patterns compared with cortical cells compared with sinusoidal gratings," *Journ. of Physiology*, vol. 204, pp. 283-298, 1969.

6. Campbell, F. and Green, D., "Optical and retina factors affecting visual resolution," *Journ. of Physiology*, vol. 181, pp. 576-593, 1965.

7. Campbell, F. and Kulikowski, J., "Orientation selectivity of the human visual system," *Journal of Physilogy*, vol. 197, pp. 437-441, 1966.

8. Campbell, F. and Robson, J., "Application of Fourier analysis to the visibility of gratings," *Journ. of Physiology*, vol. 197, pp. 551-566, 1968.

9. Crochiere, R. C. and Rabiner, L. R., "Interpolation and decimation in signal processing," *Proceeding on the IEEE*, vol. 69, March 1981.

10. Crowley, J., "A representation for visual information," *Tech. Rep. CMU-RI-TR-82-7*, Robotic Inst. Carnegie-Mellon Univ., 1987.

11. Daubechies, I., "Orthonormal bases of compactly supported wavelets," *To appear in IEEE Trans. on Information Theory*, Bell lab., 1987.

12. Daubechies, I., "The wavelet transform, time-frequency localization and signal analysis," *To appear in Communications in Pure and Applied Mathematics*, 1988.

13. Daugmann, J. G., "Two-dimensional spectral anlaysis of cortical receptive field profile," *Vision Research*, vol. 20, pp. 847-856 , 1980.

14. Daugmann, J. G., "Six formal properties of two dimensional anisotropic visual filter. Structural principles and frequency / orientation selectivity," *IEEE trans on system man and cybernetics*, vol. smc-13, Sept. 1983.

15. Esteban, D. and Galand, C., "Applications of quadrature mirror filters to split band voice coding schemes," *Proc. ICASSP*, May 1977.

16. Federbush, P., "Quantum field theory in ninety minutes," *Bull. Am. Math. Soc.*, 1987.

17. Gabor, D., "Theory of communication," *J. Inst. Electr. Eng.*, vol. 93, pp. 429-457, London, 1946.

18. Geman, D. and Geman, S., "Stochastics relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Trans. on Pattern Annal. and Mach. Intell.*, vol. 6, pp. 721-741, 1984.

19. Georgeson, M., "Mechanisms of visual image processing: studies of pattern interaction and selective channels in human vision," in *Ph.D. diss.*, Univ. of Sussex, Brighton England, 1975.

20. Georgeson, M., "Spatial Fourier analysis and human vision," in *Tutorial essays in psychology, a guide to recent advances*, ed. N. Sutherland, vol. 2, Lawrence Erlbaum Associates, Hillsdale, 1979.

21. Grimson, W., "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Analys. Machine Intell.*, vol. 7, pp. 17-34, Jan. 1985.

22. Grossmann, A. and Morlet, J., "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math.*, vol. 15, pp. 723-736, 1984.

23. Grossmann, A., Morlet, J., and Paul, T., "Transforms associated to square integrable group representations," *int. Journ. Math. Phys.*, vol. 26, pp. 2473-2479, 1986.

24. Hall, E., Rouge, J., and Wong, R., "Hierarchical search for image matching," *Proc. Conf. on Decision and Control*, pp. 791-796, 1976.

25. Hubel, D. and Wiesel, T., "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J.physiol*, vol160, 1962.

26. Hummel, R., "Representations based on zero-crossings in scale-space," Tech. Rep. 225, Courant Inst., Dept. Computer Sc., June, 1986.

27. Hummel, R. and Moniot, R., "A network approach to reconstruction from zero-crossings," *Proc. of IEEE Workshop on computer vision*, Dec. 1987.

28. Jaffard, S. and Meyer, Y., "Bases d'ondelettes dans des ouverts de Rn," *Journ. de Mathematiques pures et appliquees*, 1987.

29. Klauder, J. and Skagerstam, B., in *Coherent states*, World Scientific, Singapore, 1985.

30. Koenderink, J., "The structure of images," *Biological Cybernetics*, Springer Verlag, 1984.

31. Kronland-Martinet, R., Morlet, J., and Grossmann, A., "Analysis of sound patterns through wavelet transform," *International Journal on Pattern Analysis and Artificial Intelligence*, Jan. 1987.

32. Kulikowski, J. and King-Smith, P., "Orientation selectivity of grating and line detectors in human vision," *Vision Research*, vol. 13, pp. 1455-1478, 1973.

33. Lemarie, P.G., "Ondelettes a localisation exponentielles," *Journ. de Math. Pures et Appl.*. to be published

34. Lemarie, P. G. and Y., Meyer, "Ondelettes et bases Hilbertiennes," *Revista Matematica Ibero Americana*, vol. 2, 1986.

35. Levine, M. D., *Vision in man and machine*, Mc Graw Hill, 1985.

36. Maffei, L. and Fiorentini, A., "The unresponsive regions of visula cortical receptive fields," *Vision Research*, vol. 16, pp. 1131-1139, 1976.

37. Maffei, L., Morrone, C., Pirchio, M., and Sandini, G., *J. of Physiology*, vol. 296, pp. 24-47, 1979.

38. Mallat, S., "A theory for multiresolution signal decomposition : the wavelet representation," *To appear in IEEE Trans. on Pattern Analysis and Machine Intelligence.*, Tech. Rep. MS-CIS-87-22 U. of Penn., 1987.

39. Mallat, S., "Multiresolution approximation and wavelet orthonormal bases of L2," *To be published in Transactions of the Americain Mathematical Society.*, Tech. Rep. MS-CIS-87-87 U. of Penn., Aug. 1987.

40. Mallat, S., "Dyadic wavelets energy zero-crossings," *To appear as an invited paper in IEEE Trans. on Information Theory*, Tech. Rep. MS-CIS-88-30 U. of Penn., 1988.

41. Marr, D. and Hildreth, E., "Theory of edge detection," *Proc. of the Royal Society of London*, vol. 207, pp. 187-217, 1980.

42. Marr, D. and Poggio, T., "A theory of human stereo vision," *Proc. Royal Soc. London*, vol. B 204, pp. 301-328, 1979.

43. Meyer, Y., "Ondelettes et fonctions splines," *Seminaire Equations aux Derivees Partielles*, Ecole Polytechnique, Paris, France, Dec. 1986.

44. Meyer, Y., in *Ondelettes et Operateurs*, Hermann, 1988.

45. Meyer, Y., "Principe d'incertitude, bases hilbertiennes et algebres d'operateurs," *Bourbaki seminar*, 1985-86, no 662.

46. Nachmias, J. and Weber, A., "Discrimination of simple and complex gratings," *Vision Research*, vol. 15, pp. 217-223, 1975.

47. Paul, T., *Affine coherent states and the radial Schrodinger equation. Radial harmonic oscillator and hydrogen atom*, To be published.

48. Pollen, D. A. and Ronner, S. F., "Visual corical neurons as localized spatial frequency filter," *IEEE trans on systems man and cybernetics*, vol. smc-13, p. september 1983.

49. Rosenfeld, A., *Multiresolution image processing and analysis*, Springer Verlag, 1982.

50. Rosenfeld, A. and Thurston, M., "Edge and curve detection for visual scene analysis," *IEEE trans Comput*, 1971 vol C-20.

51. Rosenfeld, A. and Vanderburg, G. J., "Coarse-fine template matching," *IEEE Trans. Systems, Man, Cybernetics*, vol. SMc-7, pp. 104-107, 1977.

52. Sanz, J. and Huang, T., "Theorem and experiments on image reconstruction from zero-crossings," *Research report RJ5460* , IBM.

53. Smith, M. J. and Barnwell, T. P., "Exact reconstruction techniques for tree-structured subband coders," *IEEE trans. on ASSP*, vol. 34, Jun. 1986.

54. Stromberg, J., "A modified franklin system and higher-order systems of Rn as unconditional bases for Hardy spaces," *Report from Princeton University.*

55. Tanimoto, S. and Pavlidis, T., "A hierarchical data structure for image processing," *Comput. Graphics Image Process.*, vol. 4, pp. 104-119, 1975.

56. Tchamitchian, P., "Biorthogonalite et theorie des operateurs," *Revista Matematica Ibero Americana*, vol. 2, 1986.

57. Valois, K. De, Valois, R. De, and Yund, E., *J. of Physiology*, vol. 291, pp. 483-505, 1979.

58. Watson, A., "Efficiency of a model human image code," *Journ. Opt. Soc. of Amer.*, vol. 4, pp. 2401-2417, Dec. 1987.

59. Webster, M. and Valois, R. De, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," *J. Opt. Soc. Am.*, July 1985.

60. Witkin, A., "Scale space filtering," *Proc. Int. Joint Conf. Artificial Intell.* , 1983.

61. Woods, J. W. and O'Neil, S. D., "Subband coding of images," *IEEE trans. on ASSP*, vol. 34, Oct. 1986.

62. Zeevi, Y. and Rotem, D., "Image reconstruction from zero-crossings," *IEEE ASSP*, vol. 34, pp. 1269-1277, 1986.